



**Universidad Nacional Autónoma de México**  
Programa de Maestría y Doctorado en Pedagogía  
Facultad de Filosofía y Letras

**EVALUACIÓN A GRAN ESCALA DEL EGRESO DE LA LICENCIATURA: HACIA UN MODELO  
COMPRESIVO**

**Tesis**  
**Que para obtener el grado de Doctor en Pedagogía**

**Presenta:**  
José Alfonso Jiménez Moreno

**Tutor:**  
Dr. Mario Rueda Beltrán  
Instituto de Investigaciones Sobre la Universidad y la Educación

**Comité tutorial:**  
Dr. Alejandro Canales Sánchez  
Instituto de Investigaciones Sobre la Universidad y la Educación  
Dr. Alejandro Márquez Jiménez  
Instituto de Investigaciones Sobre la Universidad y la Educación  
Dr. Héctor Alfonso Vera Martínez  
Instituto de Investigaciones Sobre la Universidad y la Educación  
Dra. Gabriela de la Cruz Flores  
Instituto de Investigaciones Sobre la Universidad y la Educación

**México, abril 2016**



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Para los grandes amores en mi vida: Sayid, Uriel y Bruno,  
su existencia da sentido a la mía.

Para Andrea. Tu risa ilumina cualquier oscuridad.

## Agradecimientos

Gracias a mis hijos Sayid, Uriel y Bruno por motivarme a siempre buscar opciones de crecimiento, su bienestar es mi motor de vida personal y profesional. Gracias a ustedes diariamente aprendo a ser padre. Seguiré esforzándome para que algún día pueda hacer que se sientan tan orgullosos de mí como yo siempre me he sentido orgulloso de ustedes.

Gracias a Andrea, por nunca dejar de creer en mí, amarme de forma incondicional y siempre tener tu mano dispuesta para ayudar a levantarme.

Gracias a mi madre por tu motivación y amor a lo largo de toda mi vida. A mis hermanas Julieta y Lucía, por siempre estar al pendiente de mi bienestar. A mi padre por sentirse orgulloso de mí.

Gracias a Jacqueline, por los años de bienestar y apoyo cuando más lo necesité, así como por la madurez para que sigamos caminando.

A mi Tutor, el Dr. Mario Rueda, que me orientó en lo académico y en lo personal, siempre preocupado por mi formación. Al Dr. Alejandro Márquez, por confiar en mí y apoyarme tanto. Gracias al Dr. Alejandro Canales, por una retroalimentación pertinente de una forma tan cordial. Al Dr. Héctor Vera por ampliar mi campo de conocimiento. A la Dra. Gabriela de la Cruz, por una retroalimentación tan profunda sin perder la mirada constructiva.

Gracias a la Mtra. Lilia Jiménez Mejía de la FCPyS, con quien he trabajado de forma maravillosa. A la Dra. Ofelia Contreras, por siempre confiar en mi desempeño.

Al Ceneval y a las universidades con los que compartí tantos años y experiencias, así como a los profesores y amigos, como Luis Antonio y Leticia, que ahí conocí y que me regalaron tanta alegría. Gracias Ceneval por haberme brindado experiencia, satisfacciones y el interés por mejorar la evaluación y la educación.

A los CIEES y a los compañeros que ahí laboran, que me abrieron las puertas al mundo de la educación superior. Al INEE, que en mi estadía en él desarrollé este trabajo. A la UNAM, por la formación que me dio y las experiencias laborales que he vivido en ella.

Al programa de posgrado en pedagogía y a sus docentes por abrirme sus puertas y modificar mi visión sobre la educación.

*“La grandeza del hombre está en ser un puente y no una meta: lo que en el hombre se puede amar es que es un tránsito y un ocaso”*

Nietzsche. Así habló Zaratustra

# Índice

<b>Introducción.....</b>	<b>6</b>
<b>1. Modelos de evaluación educativa.....</b>	<b>12</b>
Modelo centrado en objetivos.....	14
Método científico de la evaluación.....	17
Modelo CIPP.....	19
La evaluación iluminativa.....	20
La evaluación centrada en el consumidor.....	22
Evaluación comprensiva.....	24
Reflexiones sobre los usos de los modelos de evaluación.....	30
<b>2. Evaluación a gran escala.....</b>	<b>34</b>
Antecedentes y conceptualización.....	34
La evaluación a gran escala como elemento de rendición de cuentas.....	38
Evaluación a gran escala en México.....	47
Evaluación a gran escala del egreso del nivel superior.....	54
<b>3. Descripción del problema y metodología.....</b>	<b>66</b>
<b>4. Análisis de caso: EGEL.....</b>	<b>78</b>
A manera de introducción: ¿qué es el EGEL?.....	78
Inicios. ¿Cuáles fueron los acontecimientos y/o necesidades que llevaron a la ANUIES a la creación de los EGEL?.....	80
Evolución. ¿Cuál es el estatus actual de los EGEL y cuáles son los aspectos sociales y académicos que han favorecido su desarrollo?.....	86
Concepción teórica del EGEL. ¿Cuál es el modelo que sustenta la actividad evaluativa del EGEL?.....	92
Metodología para su elaboración.....	99
Elaboración de reactivos. ¿Cuál es el procedimiento que se sigue para el desarrollo de reactivos?.....	107
Proceso de aplicación. ¿Cómo se desarrolla el proceso de aplicación?.....	109
Proceso de definición de puntos de corte y calificación. ¿Qué metodología se utiliza para la definición de puntos de corte y emisión de calificación?.....	110
Emisión de resultados. ¿Cuáles son los elementos de decisión que soportan la forma de presentar resultados?.....	128
A manera de cierre.....	136
<b>5. Propuesta de evaluación a gran escala del egreso de licenciatura bajo un enfoque comprensivo.....</b>	<b>139</b>
Epistemología de la evaluación a gran escala del egreso de la licenciatura.....	139

Metodología.....	147
Proceso de recopilación de información.....	173
Análisis de la información .....	175
Emisión de resultados .....	185
<b>Conclusiones .....</b>	<b>191</b>
<b>Referencias .....</b>	<b>200</b>

## Introducción

La educación superior en México, enfocada en la formación tecnológica y de profesionistas, ha dado un enorme crecimiento desde finales del Siglo XX. El Sistema de Educación Superior (SES) mexicano es complejo y heterogéneo debido a su tamaño y a las particularidades que lo integran. Existe una gran diversidad de tipos de instituciones de educación superior, se subdivide en: Subsistema de Universidades Públicas Federales, Subsistema de Universidades Públicas Estatales, Subsistema de Educación Tecnológica, Universidades Tecnológicas Públicas, Universidades Politécnicas Públicas, Subsistema de Universidades Públicas Interculturales, Instituciones Particulares, Instituciones de Formación Docente (Escuelas Normales), así como el Subsistema de Centros Públicos de Investigación (Cruz y Cruz, 2008); adicionalmente, cada entidad del país cuenta con estructuras gubernamentales y educativas propias que incrementan la diversidad de este sistema. En esencia, la formación profesional en nuestro país se caracteriza por una enorme diversidad de tipos, naturaleza, forma de sostenimiento y régimen jurídico (Rueda, 2008).

La complejidad y la heterogeneidad que distinguen a la educación superior mexicana han tenido como consecuencia una serie de problemáticas, desde la atención a la alta demanda (Tuirán, 2011), así como el problema de calidad que ofrece cada una de ellas (Rueda, 2008). Por supuesto, es de esperarse que con una diversidad de tipos de instituciones como las que presenta el SES, las características y la calidad de la formación técnica y profesional de cada una de las disciplinas difiera de gran manera entre cada una de las instituciones y escuelas. Frente a ello, y con fines de aseguramiento de la calidad de la formación superior mediante una intención de establecer una organización homogénea en su gestión, la evaluación de la educación superior se institucionalizó con el Programa para la modernización educativa 1989-1994 del Gobierno Federal; En este programa se instituyó como prioritaria la evaluación interna y externa de los programas

educativos, así como la creación de una instancia que integrara el proceso de nacional de evaluación de la educación superior. Para ello, la Coordinación Nacional para la Planeación de la Educación Superior (CONPES) creó en 1989 la Comisión Nacional de Evaluación de la Educación Superior (CONAEVA), así como los Comités Interinstitucionales para la Evaluación de la Educación Superior (CIEES) en 1991 y el Centro Nacional de Evaluación para la Educación Superior (Ceneval) en 1994 (Cruz y Cruz, 1998).

La consideración de agentes externos como medio de organización y regulación de la calidad de la formación académica superior es, por supuesto, un reflejo de que el registro oficial de un programa educativo a nivel técnico o profesional no implica necesariamente el aseguramiento de calidad educativa para los intereses del Estado. Es entonces que el tema de la evaluación de los procesos educativos ha cobrado particular importancia dentro de la política educativa mexicana de las últimas décadas.

De manera específica, la forma en que tradicionalmente se ha evaluado la calidad de una formación académica a nivel superior ha sido considerando la identificación del desempeño de los estudiantes en función de una disciplina en específico. En particular en el caso de México, el inicio de la evaluación de los egresados de nivel licenciatura ha sido una herramienta importante para el avance del sistema de evaluación de la educación superior, lo cual dio origen a instituciones clave para el entendimiento de la dinámica de las universidades encargadas de la formación de profesionistas. Este tipo de evaluación se constituyó bajo una metodología comúnmente conocida como *a gran escala*, la cual facilita la sistematización de la formación académica de nivel superior; aunque, por supuesto, siempre ha existido la valoración que las instituciones educativas realizaban respecto a los logros alcanzados por sus estudiantes con fines de obtener un reconocimiento legal para desempeñar su profesión. Una evaluación de forma tan directa sobre los estudiantes

permite, sin duda, la verificación de los avances y logros que los estudiantes desarrollaron durante su formación profesional, aunque el objetivo de generar una evaluación de este tipo bajo una metodología a gran escala representaba para las instituciones mexicanas la posibilidad de comparar programas académicos.

El desarrollo de evaluaciones de resultados educativos mediante un acercamiento a gran escala permitió el establecimiento de formas metodológicas que innovaron el campo de la medición de los aprendizajes, igualmente, apoyó a sentar las bases de una cultura de la evaluación educativa dentro del ámbito de la educación superior en México. Durante más de dos décadas este modelo favoreció la cultura de la rendición de cuentas, así como la intención de sistematización de los conocimientos esperados dentro de diversas formaciones profesionales, los cuales son elementos que tradicionalmente se consideran valiosos para el análisis de la calidad de la educación. Sin embargo, en la actualidad los diversos enfoques pedagógicos mediante los cuales se desarrollan procesos de evaluación en la educación han mantenido una evolución que ha traspasado la metodología tradicional centrada únicamente en la consideración de la medición de logros de estudiantes fuertemente centrados en el dominio de contenidos disciplinares. Al respecto, como se analizará más adelante, la innovación del modelo de evaluación a gran escala de los egresados de nivel superior se ha quedado al margen, a pesar de la enorme relevancia política que tiene para las universidades en el rumbo de su vida institucional.

Adicionalmente, los resultados que los egresados universitarios obtienen en las llamadas *pruebas a gran escala* que se utilizan para valorar la calidad de su formación también suelen utilizarse como medio de reconocimientos, o bien, se consideran como una condición a su egreso o titulación. Como se analizará más adelante, estas prácticas también han sido nocivas y han mermado la innovación pedagógica en este tipo de evaluación.

El presente trabajo representa un ejercicio de verificar el estado actual de la evaluación a gran escala, y, en particular, de la enfocada a la evaluación del egreso del nivel profesional, ello mediante una revisión de la epistemología, metodología y técnica de diversas prácticas internacionales, pero en particular de las nacionales. Esta revisión da pauta a la generación de un estudio de caso en particular: el *Examen General para el Egreso de la Licenciatura* desarrollado por el Centro Nacional de Evaluación para la Educación Superior, el cual es uno de los ejercicios pioneros en México en lo que respecta a este tipo de evaluación. El análisis desarrollado favorece la comprensión de las dinámicas sociales y políticas que generaron la necesidad de establecer una evaluación a gran escala del egreso del nivel profesional. Como en cada acción científica, los primeros ejercicios suelen ser los más criticados, pero es gracias a ellos que se pueden sentar las bases de propuestas que permitan afrontar nuevas dinámicas y renovar técnicas para hacerlas más acordes con las necesidades académicas y sociales que se viven conforme estas evolucionan. Bajo esa lógica, el análisis de caso del Examen General para el Egreso de la Licenciatura facilita el desarrollo de una propuesta de evaluación a gran escala del egreso del nivel licenciatura bajo un enfoque comprensivo, el cual parece responder de mejor manera a las formas de trabajo educativo que se viven en estos inicios del siglo XXI y permite también optimizar algunas de las dificultades que la forma tradicional de este tipo de evaluación presenta.

El interés de plantear una propuesta alterna para desarrollar una evaluación a gran escala enfocada al egreso del nivel licenciatura radica principalmente en la identificación de las inconsistencias epistemológicas de la forma actual de realizar este tipo de evaluación. Este interés se generó en función de la experiencia que he tenido en proyectos de impacto nacional dentro del Centro Nacional de Evaluación para la Educación Superior como coordinador del Examen General para el Egreso de la Licenciatura en Comercio/Negocios Internacionales de 2009 al 2011, así como del Examen General para el Egreso de la Licenciatura en Turismo del 2009 al 2013 y como

coordinador iniciador del Examen General para el Egreso de la Licenciatura en Gastronomía durante 2012 y 2013; así como subdirector de pruebas internacionales dentro del Instituto Nacional para la Evaluación de la Educación desde el año 2013, trabajando en proyectos como el *Programme for the International Student Assessment (PISA)*, *International Civic and Citizenship Education Study (ICCS)* y los estudios generados por el Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE).

El presente trabajo se desarrolla mediante cinco capítulos. En el Capítulo 1, se describen los diferentes modelos de evaluación en educación que son mencionados con mayor frecuencia en la literatura, con el fin de mostrar la diversidad existente de posturas evaluativas. El Capítulo 2 presenta el estado actual de la evaluación a gran escala, y, en particular, aquella enfocada al egreso de licenciatura, indicando elementos políticos y académicos que dieron pauta a su origen, así como precisando de forma general la metodología utilizada. Con base en el encuadre que otorgan estos dos primeros apartados, el Capítulo 3 aborda de forma más detallada la argumentación y metodología del desarrollo de la propuesta de evaluación a gran escala del egreso de la licenciatura de forma comprensiva. El Capítulo 4 se enfoca en el estudio de caso del Examen General para el Egreso de la Licenciatura considerando como base de análisis la manera en que se ha realizado dicha actividad en nuestro país, ello mediante la verificación de documentos oficiales y rescatando también mi propia experiencia en estos proyectos. Por su parte, en el Capítulo 5 se aborda una propuesta para desarrollar una evaluación a gran escala del egreso de la licenciatura mediante un enfoque comprensivo, haciendo énfasis en la coherencia epistemológica dentro de sus diversas fases, así como resaltando la importancia de atención a la diversidad, así como la participación de diversos agentes involucrados con el egreso de la licenciatura. Por último, en las Conclusiones se resaltan los elementos más relevantes de la propuesta desarrollada, así como una reflexión sobre la viabilidad de la propuesta, considerando

la función académica y política que le ha dado cabida a la evaluación del egreso a gran escala en el SES actual.

## 1. Modelos de evaluación educativa

La evaluación en educación ha tenido un auge importante a nivel mundial en las últimas décadas (Tiana, 2012a), ya que es una actividad a la cual se le ha reconocido su valor como elemento importante de retroalimentación de avances educativos, tanto dentro del aula, como a un nivel macro, tal es el caso de los sistemas educativos. La evaluación de la educación se ha profesionalizado desde mediados del siglo XX debido a la creciente necesidad social de contar con metodologías cada vez más amplias y precisas que faciliten la comprensión de fenómenos del ámbito de la educación con fines de retroalimentación y mejora de las prácticas educativas a diversos niveles de escolaridad, por medio del establecimiento de políticas educativas, hasta la mejora de estrategias de enseñanza dentro del aula, tanto en México, como en el resto de mundo (Martínez, 2009). Desde una postura académica, la evaluación en educación se ha usado con dos propósitos fundamentales, el primero de ellos, el llamado *formativo*, se enfoca principalmente a analizar diversos aspectos para poder realizar una retroalimentación con fines de mejora del evaluado (Casanova, 1998); normalmente es llamada de *proceso*, sin embargo, el objetivo formativo o de retroalimentación no debe estar supeditado a un momento en particular. Por su parte, el objetivo *sumativo* es el que se relaciona con toma de decisiones –tradicionalmente de alto impacto– en función de la obtención de una valoración en particular, las cuales no tienen la intención precisa de retroalimentar al evaluado; los ejercicios más comunes de la evaluación sumativa es el ingreso o egreso a nivel educativo, o bien, la certificación de conocimientos o la acreditación de programas educativos.

El interés respecto al conocimiento del impacto que tienen las políticas y los programas educativos en esta materia ha tenido una larga historia, incluso puede remontarse a la Ilustración (Chiva, Perales y Pérez, 2009), pero es a partir del siglo XIX que comenzó el uso de la estadística como una herramienta importante de decisión política, lo cual tuvo una enorme influencia sobre procesos de evaluación, siendo así coincidente con la tendencia positivista en ciencias sociales característica de esa época. Después de la Segunda Guerra Mundial, la búsqueda de diferencias individuales a través de la medición de rasgos psicológicos, como elemento derivado de esta tendencia metodológica positivista, así como la búsqueda de evidencia empírica de procesos de aprendizaje tuvo un auge importante, de forma particular en los Estados Unidos. El uso de tests estandarizados permitía contar con una medida de evaluación de programas educativos bajo una

política de búsqueda de rendición de cuentas, ello debido a la fuerte inversión que el gobierno de dicho país había hecho sobre programas de mejora educativa como consecuencia de los problemas acarreados por la Gran depresión. El enorme interés del gobierno estadounidense de tener una educación de calidad promovió que los académicos centraran su atención en buscar maneras de verificar que los programas educativos eran exitosos (Chiva, Perales y Pérez, 2009; Stufflebeam y Shinkfield, 1993). Los sistemas políticos basados en la rendición de cuentas, los cuales pueden basarse en el cumplimiento de reglamentos, los que expresan el apego a normas profesionales y los que son impulsados por la medición de resultados, comenzaron a interesarse en la observación del logro educativo para valorar que las escuelas públicas funcionaran de la manera esperada para lograr una superación económica nacional (Anderson, 2005).

Bajo este contexto, Ralph Tyler en su obra *Principios básicos del currículo* de 1950, desarrolló una propuesta en la que manifestaba que la evaluación debía observar la congruencia entre los objetivos de un programa educativo y los logros obtenidos mediante su aplicación. El establecimiento de objetivos educativos de forma mensurable permitía la implementación de intervenciones y evaluaciones educativas centradas en elementos observables, facilitando así su implementación y su posterior evaluación. De esta forma, la evaluación de la educación comenzó a entenderse como un proceso que determina hasta qué punto los objetivos de un programa educativo se habían alcanzado, y, dada la tendencia empirista y estadística vivida en esos momentos, la verificación de avances educativos se podía obtener a través de la observación de los logros o aprovechamiento de los estudiantes en función de objetivos mensurables previamente establecidos.

La propuesta de Tyler fue considerada una opción viable y relevante para la identificación de logros académicos de los diversos programas educativos durante diversos años y se le comenzó a considerar un modelo de evaluación conforme inició el desarrollo de otras perspectivas evaluativas en educación. Actualmente hay una creciente diversidad de perspectivas y metodologías que favorecen el proceso de retroalimentación de acciones educativas adicionales a la propuesta por Tyler. Los diferentes modelos que continuaron a los planteamientos evaluativos de Tyler han permitido fortalecer la metodología de evaluación educativa, en particular la relacionada con la evaluación de programas educativos. Sin embargo, como se analizará en capítulos posteriores, la evaluación a gran escala ha mantenido una línea de trabajo enfocada

particularmente en los postulados tylerianos. Si bien, como se analizará, en este tipo de evaluación no suelen realizarse trabajos bajo un modelo distinto, vale realizar una aproximación a los diversos modelos de evaluación educativa y cómo éstos resultan valiosos para efectos de considerar la posibilidad de desarrollar la evaluación del aprendizaje bajo un enfoque diferente.

A continuación se describen las principales características de los modelos de evaluación mayormente utilizados. La descripción de los elementos característicos de las principales perspectivas en evaluación educativa permitirá identificar el abanico de posturas existente para poder retroalimentar los procesos y logros educativos, con fines de comprender la complejidad que caracteriza a cada una de las posturas y los elementos que cada una de ellas implica.

### **Modelo centrado en objetivos**

El llamado Modelo centrado en objetivos fue propuesto por Tyler en la década de los cincuenta del siglo XX con el fin de determinar si los programas educativos estadounidenses tenían un efecto respecto al rendimiento de los alumnos. El proceso de evaluación propuesto por Tyler implica inicialmente establecer metas u objetivos educativos por alcanzar dentro de un programa o intervención educativa (Stufflebeam y Shinkfield, 1993), este no es un proceso sencillo, pero marca la línea primaria para el desarrollo de la propuesta educativa, así como de su evaluación. Cuando esto se logra lo consecuente es ordenar los objetivos en amplias clasificaciones, lo cual permite su organización, además de establecer las dimensiones que consideran el espectro de evaluación. Posteriormente es necesario definir los objetivos en términos de comportamiento, lo cual permite su mensurabilidad, dado que se operacionalizan con fines de medición. Ahora, con fines de complementar esta operacionalización, se deben establecer situaciones y condiciones según las cuales puede ser demostrada la consecución de objetivos, lo cual permite estandarizar su implementación (elemento que posteriormente facilita su medición y registro). Contando ya con el insumo mensurable de lo que se espera de la formación educativa es posible escoger o desarrollar las apropiadas medidas técnicas de intervención o aplicación, posteriormente, recopilar los datos de trabajo (en el caso de los programas educativos, deben referirse al trabajo de los estudiantes) y finalmente comparar los datos obtenidos con los objetivos de comportamiento inicialmente establecidos. Es este último punto lo que cierra el proceso y permite

valorar los avances a la luz de la sistematización planeada a lo largo del proceso, siendo entonces los resultados una especie de *espejo* respecto a los objetivos planteados.

Esta perspectiva –que le valió a Tyler el título del *padre de la evaluación educativa*– implicaba entonces considerar al aprendizaje como una modificación de conducta resultante de un programa educativo, y, por tal, involucraba la definición de objetivos curriculares en términos de comportamiento observable, establecer situaciones y condiciones según las cuales puede ser demostrada la consecución de objetivos, recopilar los datos de trabajo para su comparación con lo establecido inicialmente, ello con el fin de retroalimentar el alcance de los programas educativos.

La propuesta de Tyler, a pesar de estar centrada en la observación de logros cuantificables y medibles, se pensó originalmente con fines de retroalimentación y no como herramienta de decisiones sumativas que respondieran a la rendición de cuentas (Stufflebeam y Shinkfield, 1993), o bien, al otorgamiento de recursos o para el establecimiento de un ranking de desempeño; irónicamente, en la práctica se ha utilizado únicamente con el objeto de valorar los logros para establecer fines sumativos, como sucede con evaluaciones a gran escala de ingreso a la educación media superior y superior. Para ahondar más en ello, el Capítulo 2 se presenta una descripción de diferentes tipos de evaluaciones centradas en este modelo que se utilizan con objetivos sumativos. Incluso, es esta perspectiva formativa, o de retroalimentación hacia la mejora, lo que marcó un parte aguas dentro de la tradición de medición psicológica que le precedió a Tyler, ya que si bien la medición era un sinónimo de la evaluación, de ahora en adelante ya se pretendía también una mejora de lo analizado.

Como se mencionó previamente, dado que esta perspectiva concebía al aprendizaje como un cambio conductual producto de los programas educativos (ello con fuerte influencia empírico-analítica en boga por la búsqueda constante de evidencias observables para la toma de decisiones), los objetivos educativos siempre debían estar establecidos en términos operativos, ya que de esa manera podían verificarse una vez finalizada la intervención educativa. Un panorama de estas características representó entonces una gran oportunidad para el desarrollo de investigaciones con fines evaluativos, dada su relativa facilidad de implementación considerando únicamente la medición de rasgos psicológicos (en particular, el aprendizaje), lo cual propició un importante desarrollo técnico en el uso de métodos cuantitativos. Sin embargo, a pesar del avance que este modelo representó, también se considera una de sus grandes desventajas, ya que los

elementos relacionados con el aprendizaje que no necesariamente son operacionalizables (como la reflexión sobre un tema o el juicio individual) quedan fuera de toda valoración para fines de retroalimentación del objeto de evaluación<sup>1</sup>.

La tendencia centrada en medición y estadística del modelo tyleriano con fines de observación y juicio del cumplimiento de objetivos permitió un avance técnico importante respecto a las formas de medir el aprendizaje como evidencia de la calidad de los programas educativos, sin embargo, ya en la época de los sesenta, Cronbach (Chiva, Perales y Pérez, 2009) planteaba la necesidad de ampliar los métodos de la evaluación educativa. En su texto *Course improvement through evaluation*, Cronbach propone que es necesario desarrollar nuevos métodos de evaluación educativa, así como hacer estudios experimentales o a pequeña escala con fines de aseguramiento de la calidad de los procesos de investigación centrados en definir el valor o mérito de algo. Este punto es lo que según Cronbach pudiera ayudar a la legitimidad científica a la evaluación. Cronbach también asegura que la evaluación no solo debe hacerse con tests psicológicos, sino con cuestionarios, entrevistas, etcétera, lo cual permitirá ampliar la diversidad de formas de acercamiento a los fenómenos y evitar limitar la forma de su observación. De acuerdo con su postura, la evaluación debe hacerse a lo largo de todo el proceso educativo, no solo para atender resultados, ya que de esa forma se puede orientar de mejor manera la práctica educativa. Asimismo, bajo esta perspectiva, al desarrollar una evaluación debe pensarse en criterios absolutos en los estándares de comparación, de tal forma que la reflexión sobre la calidad de lo evaluado no sea relativo al juicio de quien desarrolla los estándares, sino en función de los criterios mismos. Por otra parte, Cronbach sugiere que la evaluación debe centrarse más en estudios estructurales que en estudios comparativos, ya que, al igual que el punto previo, ello facilita que los estándares bajo los cuales se realiza una evaluación no dependan de un momento en particular bajo el cual se desarrollan. Por último, afirma que la evaluación no debe basarse en los objetivos, sino en los resultados que pretende informar. Sin duda, las críticas de Cronbach representan una innovación para la evaluación del momento, al considerar la necesidad de ampliar las formas metodológicas en su implementación, pero, sobre todo, una de las recomendaciones más importantes de esta perspectiva es la invitación a virar el enfoque de la

---

<sup>1</sup> Los términos “objeto de evaluación” y “evaluado” se usan de forma indistinta en este trabajo para referir al elemento educativo que se encuentra en proceso de investigación con fines evaluativos o de juicio; por ejemplo, el aprendizaje de alumnos, programas educativos, materiales educativos, docencia, sistemas educativos, etc.

evaluación centrada en objetivos hacia la presentación de toda una gama de resultados y presentar así una retroalimentación mucho más amplia de lo que la perspectiva tyleriana puede ofertar. Como se puede observar, las críticas de Cronbach se enfocaron fuertemente en la reflexión sobre la metodología en sí misma, con el fin de que pudiera valerse de elementos científicamente aceptados para llevarla a su reconocimiento social como una práctica con solidez técnica suficiente como para valerse de ella en la toma de decisiones.

Por otra parte, en años más recientes, hay autores como Antúnez y Aranguren (2004) que consideran que la evaluación y sus diferentes modelos (entre ellos el propuesto por Tyler) tienden a una perspectiva positivista que refleja el cumplimiento de objetivos de aprendizaje como la forma de develación de verdad educativa de una persona. De acuerdo con estos autores la evaluación, entonces, debiera considerar la realidad histórica-social-cultural del evaluando, los sujetos actores, alumnos y docentes que constituyen el proceso educativo, el conocimiento interdisciplinario que puede complementar bajo diferentes perspectivas un fenómeno educativo en particular, así como la normativa educativa y otros elementos contextuales. Por supuesto, estas críticas favorecen la consolidación de la evaluación mediante un robustecimiento constante no solo de sus aspectos metodológicos (como lo hace Cronbach), sino también en lo que respecta a su condición como agente histórico-social dentro de la dinámica educativa.

Es frente a esta posibilidad de amplitud de los métodos de evaluación y de diversificación metodológica que comienza una proliferación de modelos de evaluación educativa, generados básicamente por la necesidad de retroalimentar los programas educativos en Estados Unidos y posteriormente retomados en otros países. A continuación se mencionan las características básicas de los modelos de evaluación educativa más representativos en la actualidad posteriores al modelo de Tyler y a las innovaciones de Cronbach.

### **Método científico de la evaluación**

Es un método propuesto por Edward Suchman (Stufflebeam y Shinkfield, 1993), quien partía de la idea que la evaluación debía considerar el método científico en su desarrollo, ya que ello llevaría al evaluador a obtener resultados objetivos y veraces para poder emitir juicios sobre lo observado. Suchman fue el autor que comenzó a diferenciar entre investigación y evaluación, considerando que la distancia entre ambos términos radicaba en que la primera no necesariamente conllevaba a

cambios para la mejora del fenómeno analizado. Esta afirmación es relevante, debido a que comienza a establecerse de manera más formal que la evaluación se debe orientar hacia la mejora de lo observado. Si bajo el modelo de Tyler la evaluación retomó elementos propios de la investigación, y en particular, de la medición psicológica, para el establecimiento de métodos que permitirán la identificación del mérito del objeto en cuestión, para Suchman, la evaluación no era propiamente una búsqueda de la *verdad* como lo persigue la investigación, pero sí era necesario que incorporara elementos de diseño y metodológicos de ésta con fines de asegurar un juicio objetivo para una adecuada delimitación de un mérito en particular para un evaluando.

La perspectiva de Suchman, si bien un tanto más rígida que la expuesta por Tyler en lo que a elementos metodológicos se refiere, mantenía una línea centrada en la búsqueda objetiva de elementos operacionalizables, incluso controladas por diseños de investigación experimental, como los que se utilizan en la investigación psicológica básica. Bajo esta perspectiva, los elementos o categorías que debían caracterizar a cualquier evaluación de programas son: el esfuerzo implicado en el programa, el trabajo que lo definió, la suficiencia de este trabajo, la eficiencia obtenida, así como el proceso en general.

Adicional a considerar a la evaluación como un proceso que debe cuidar su diseño como lo hace una investigación, Suchman comienza ya a visualizar la posibilidad de incluir más elementos que solo el cumplimiento de objetivos para poder definir el mérito o valía de un fenómeno educativo, llegando a considerar elementos de esfuerzo y proceso, aunque siempre con el peso implícito de los objetivos; ejemplo de ello es la búsqueda de la evaluación de la eficiencia, la cual siempre estará ligada a lo que se pretende alcanzar, independientemente de que también tome en cuenta los tiempos y los recursos con los que se cuenta. A pesar de ello, la consideración de elementos alternos a la mera observación del cumplimiento de objetivos representó un avance relevante para el desarrollo de los modelos de evaluación.

Ahora, si bien Suchman manifiesta que no se puede alcanzar del todo un conocimiento completo sobre un objeto de evaluación aun considerando un diseño de investigación robusto, es importante utilizar un método estricto cuando se pretenda desarrollar una evaluación, con el fin de retroalimentar los logros finales y elementos intermedios que hayan influido en el cumplimiento de objetivos educativos. La gran ventaja de este modelo es la búsqueda para identificar elementos característicos del objeto de evaluación, sin embargo, al igual que el modelo

de Tyler, la preconcepción de variables que influyen en lo que se pretende observar deja fuera la posibilidad de analizar elementos que no se hayan planeado con anterioridad, dejando así nula la opción de inclusión de elementos educativos que no hayan sido previstos en el esquema prediseñado.

## Modelo CIPP

Uno de los modelos más utilizados en evaluación educativa es el llamado CIPP propuesto por Stufflebeam (Stufflebeam y Shinkfield, 1993). Las siglas CIPP hacen referencia a los elementos que el modelo considera que deben ser observados para emitir un juicio de valor sobre un programa educativo, a saber: Contexto, Insumos, Procesos, y Productos.

Stufflebeam presenta esta iniciativa siguiendo una línea de desarrollo de la metodología evaluativa que le precede, al considerar elementos adicionales a los objetivos, como ya Suchman apuntaba, pero siendo consciente de la complejidad de relaciones existentes dentro de un programa educativo como objeto de evaluación. Cabe mencionar que actualmente se hace uso del modelo no solo para la evaluación de programas, sino también para procesos relacionados con proyectos laborales, personal de una organización, productos y organizaciones (Stufflebeam y Shikfield, 2007, en Zhang, Zeller, Griffith, Metcalf, Williams, Shea y Misulis, 2011). Ahora, cada uno de los elementos mencionados puede representar una gran complejidad en sí mismo, dependiendo de las características del programa educativo y los objetivos que se pretendan con la evaluación (ya sean formativos o sumativos). En los elementos de la evaluación de *Contexto* deberán considerarse las necesidades de la población de interés de los programas, y si sus objetivos se relacionan con las necesidades del contexto y de los beneficiarios y su comunidad. Los *Insumos* regularmente se entienden como esos aspectos que guardan relación con el diseño educativo o pedagógico para dar respuesta a las metas planteadas, así como a las características del grupo de trabajo, recursos materiales y demás aspectos que faciliten el cumplimiento de un programa educativo y su relación con el contexto en particular. Los elementos de *Proceso* se relacionan con la identificación de cómo se está llevando a cabo el programa educativo en cuestión con el fin de retroalimentar si las actividades se han realizado conforme con lo planeado; ello implica la identificación de información cualitativa y cuantitativa que permita saber si es necesario realizar alguna modificación al programa, en virtud de que se estén alcanzando los objetivos y metas

deseadas. El último componente, *Productos*, pretende medir e identificar logros alcanzados por el programa en términos de valor, mérito, significancia y alcance de las expectativas; entre estos elementos, la evaluación del aprendizaje ha obtenido un lugar en particular como un aspecto propio de los productos por considerar para la evaluación de un programa educativo. Cada uno de los cuatro elementos de este modelo permite identificar diversos aspectos que componen un programa educativo bajo una metodología mixta (conjuntando técnicas cuantitativas y cualitativas), abarcando así una gran diversidad de cuestiones inmersas en la acción educativa que facilitan su comprensión considerando una gran diversidad de elementos dentro de la educación, los cuales abarcan la complejidad característica de los procesos educativos.

El modelo basado en la valoración de Contexto, Insumos, Procesos y Productos propuesto por Stufflebeam indica que la evaluación debe orientarse a ayudar a la toma de decisiones. Dado que el desarrollo de un programa es un proceso (en curso o concluido en el momento de la evaluación), la evaluación debe reflejar información para la toma de decisiones durante la implementación de un programa, es decir, sigue una lógica predominantemente formativa. Bajo esta lógica, “una evaluación del proceso es una comprobación continua de la realización de un plan” (Stufflebeam y Shinkfield, 1993, pág. 196); es decir, con el desarrollo de este modelo, la evaluación comienza a concebirse como la actividad que permite modificar un programa educativo y analizar la medida en que se va alcanzando lo que se ha propuesto, no solo con una mirada del logro de objetivos, sino considerando elementos relacionados con su ejecución.

Sin duda, el modelo CIPP, al considerar diversos elementos inmersos en el desarrollo de programas educativos y bajo una postura mixta en cuanto a su metodología, representa un modelo amplio que ha traído como consecuencia su uso constante de la evaluación de programas, en particular gracias a su enfoque formativo, aun cuando pueda centrar su atención en los productos o resultados.

### **La evaluación iluminativa**

Sin duda, el modelo de Tyler es la mayor influencia en evaluación educativa, por ende, ha sido el modelo de mayor crítica, aunque es eso lo que ha facilitado el avance teórico y metodológico de la evaluación en educación. Una de las mayores críticas a la propuesta de Tyler es que realiza valoraciones sobre un evaluando sin necesariamente desarrollar una explicación del porqué de los

datos observados. Es a partir de esta crítica que Parlett y Hamilton (Stufflebeam y Shinkfield, 1993) proponen la llamada *evaluación iluminativa*, la cual pretende describir a profundidad al evaluando, más que hacer uso de los números como elemento base para la emisión de un juicio de valor.

De acuerdo con Parlett y Hamilton, la educación se caracteriza por elementos que no son completamente acordes con lo que Tyler propone, en primer lugar, que la educación contiene aspectos que no necesariamente pueden ser controlados (lo cual difiere fuertemente con el esquema rígido planteado bajo el modelo centrado en objetivos). Por otra parte, el evaluador debe ser capaz de adaptarse a situaciones cambiantes, considerando que un programa no es el mismo en el proceso que se desarrolla su evaluación. Otro de los elementos es que los métodos tradicionales de evaluación suelen dejar fuera datos relevantes, dado que solo se centran en los meramente cuantificables, además de que los métodos estadísticos suelen ser menos sensibles a las particularidades de los fenómenos educativos. Por último, Parlett y Hamilton acentúan que dado que tradicionalmente se piensa en una “verdad objetiva”, se dejan de lado una diversidad de elementos que pudieran ser relevantes para la comprensión de un evaluando en particular.

Bajo esta crítica, la evaluación iluminativa implica entonces considerar dentro del proceso de evaluación al sistema de enseñanza inmerso en el evaluando, así como el medio de aprendizaje en el cual se desarrollan sus actividades, ya que cada evaluando presenta un fenómeno único en relación con su contexto y el modelo educativo que subyace en él. La evaluación iluminativa implica entonces que las decisiones técnicas del diseño y metodología de una evaluación dependerán del caso en concreto y, por ende, no se busca la generalización en su desarrollo y resultados, como sucede con la investigación educativa.

Respecto a su metodología, la principal herramienta es la observación, aunque también puede hacerse uso de cuestionarios, pero no como una herramienta única. Incluso, la evaluación iluminativa puede desarrollarse a gran escala, ya que puede contribuir a describir adecuadamente los fenómenos educativos masivos mediante terminología compartida y conceptos útiles para todos los involucrados en el proceso (Stufflebeam y Shinkfield, 1993).

La evaluación iluminativa es un modelo importante, en particular para la evaluación de programas educativos, ello debido a que su énfasis en conocer a profundidad a un evaluando mediante una metodología predominantemente cualitativa (y, en particular, etnográfica) representa un parte

aguas respecto a la metodología predominantemente cuantitativa que caracterizó a los primeros trabajos de evaluación en educación.

### La evaluación centrada en el consumidor

El modelo centrado en el consumidor fue propuesto por Scriven (Stufflebeam y Shinkfield, 1993) y entiende a la evaluación como la determinación sistemática y objetiva del valor o el mérito de algo. El énfasis de Scriven es en particular en la posibilidad de que los evaluadores tengan la capacidad de llegar a emitir juicios de valor completamente justificables, más que en la medición de rasgos o en la identificación del logro de objetivos. Ello implica que el evaluador sea una especie de sustituto del consumidor o usuario, por lo que el papel del evaluador es fungir como una “conciencia social informada” (Stufflebeam y Shinkfield, 1993; pág. 342), siempre considerando a la objetividad como elemento indiscutible de la evaluación.

Para Scriven existen dos vertientes en la evaluación, la primera es la evaluación intrínseca, la cual se centra en la valoración de las cualidades de los procesos de mediación del evaluando<sup>2</sup> (metas, estructura, metodología, actitudes, personal, etc.), y la evaluación final, la cual no se preocupa por la naturaleza del programa, sino por sus efectos sobre los clientes o audiencias involucradas con los resultados del programa, ya que considera que son éstos quienes realmente pueden valorarlo, al ser ellos los que tienen contacto con el producto de todos los esfuerzos realizados, además que están involucrados en las condiciones contextuales que determinan la mayor parte de los elementos que constituyen a un programa educativo en particular.

Una de las propuestas por las cuales se ha dado a conocer la propuesta de Scriven es debido al establecimiento de la evaluación *libre de metas*, el cual implica que el evaluador se mantenga al margen de conocer los objetivos del programa e investigar los efectos conseguidos por el programa sin este antecedente, de esa manera se pueda conocer con mayor profundidad el programa evitando un posible sesgo del evaluador cuando éste conoce previamente la intención del objeto por evaluar, aumentando así la objetividad en su labor, elemento, que como se ha observado, es de gran valía para las diversas posturas evaluativas. Por supuesto, ello representa un problema, ya que bajo la tradición tyleriana, la significación de lo evaluado se da en función de

---

<sup>2</sup> Para Scriven, el evaluando puede ser más amplio que un programa en educación, incluso su metodología puede extenderse a otros ámbitos, como a los negocios, por ejemplo.

los resultados comparados con las metas establecidas. Como solución, Scriven afirma que el evaluador debe comparar lo observado no con las metas, sino con las necesidades que identifican los consumidores del programa evaluado. Las necesidades expresadas de los consumidores del programa representan así una medida certera de lo que se espera lograr.

El proceso evaluativo bajo este modelo implica múltiples dimensiones, a saber: *Descripción*. Implica que el evaluador indague qué es lo que se debe evaluar y sus componentes. *El cliente*. No es un elemento tan sencillo de identificar, ya que quien inicia la evaluación puede no ser necesariamente el planificador del programa por evaluar. *Antecedentes y contexto*. Se debe identificar la naturaleza del evaluando, así como las expectativas que se generan alrededor de éste. *Recursos*. Se deben de considerar los recursos disponibles para la utilización del evaluando, así como para el desarrollo de la evaluación; la viabilidad estará en función de la tecnología disponible para el programa, el dinero y las consideraciones de flexibilidad para afrontar cambios en el programa. *Función*. El evaluador debe identificar qué es lo que hace el evaluando y distinguir entre lo que se supone que hace y lo que realmente realiza. *Sistema de distribución*. Corresponde al evaluador identificar: ¿Cómo es que el evaluando se da a conocer? ¿Cómo son sus usuarios? ¿Cómo se perfecciona? *El consumidor*. El evaluador deberá identificar a los consumidores, ¿quiénes son las personas afectadas de forma directa por el evaluando? ¿Quiénes son los receptores o clientes? *Necesidades y valores*. ¿Cuáles son los elementos que los afectados y potencialmente impactados identifican como valiosos o necesarios? *Normas*. El evaluador debe indagar sobre la existencia de normas preexistentes y objetivamente valoradas sobre el mérito o valor del evaluando. *Proceso*. En esta dimensión se incluyen elementos de índole administrativa, científicas e incluso éticas relacionadas con la manera en que se desarrolla el evaluando. *Resultados*. Se deben identificar los efectos producidos por el evaluando sobre la población afectada. *Posibilidad de generalización*. ¿Es posible una distribución o modificación del evaluando en otros lugares, personas, tiempos o versiones? *Costos*. Esto no implica únicamente elementos monetarios, sino también cuestiones psicológicas, o bien, directos e indirectos sobre los involucrados. *Comparaciones*. Si es posible, el evaluador debe comparar al evaluando con otros programas similares o que funjan como competencia. *Significado*. Representa una síntesis ponderada (principalmente bajo la visión del cliente) de todos los elementos previos. *Recomendaciones*. Es muy probable que quien contrata la evaluación pueda no solicitar una recomendación, sin embargo, es un elemento que no puede alejarse de la evaluación. *Informe*. Sus

características en lo que respecta a extensión, uso de lenguaje técnico y su circulación es algo que deberá ser valorado de forma cuidadosa por parte del evaluador. *Metaevaluación*. La evaluación debe ser evaluada antes de a su realización, así como antes de la emisión del informe. Los resultados deben ser utilizados con fines formativos, pero también pueden comunicarse al cliente a las audiencias interesadas.

El modelo centrado en el consumidor es un planteamiento muy completo, en el sentido que considera diversas dimensiones propias de la evaluación, además que la propuesta de evaluación libre de metas favorece una comprensión del evaluando sin la consideración de posibles prejuicios respecto al programa en función de los objetivos explícitos. Por supuesto, tomar en cuenta al consumidor o usuario final de los programas por evaluar como el eje del análisis permite sensibilizar respecto a la relevancia de evaluar los impactos y conocer las perspectivas de las personas beneficiarias de lo que se evalúe, dando así un giro importante en lo que respecta a una concepción tradicional que considera al evaluando el eje único de la evaluación y a la metodología por utilizar como el factor único de discusión.

## Evaluación comprensiva

Robert Stake es uno de los autores en modelos de evaluación de programas educativos más reconocidos, su propuesta ha sido llamada comúnmente evaluación *responsiva* o *respondente*, en virtud de que se centra en dar respuesta a las necesidades y particularidades del programa en cuestión; sin embargo, él llama a su modelo *Evaluación comprensiva* (Stake, 2006). El modelo de Stake sugiere que la evaluación necesita desarrollarse con diferentes audiencias y responder a sus necesidades evaluativas para facilitar la toma de decisiones. En este sentido, difiere de lo propuesto por Suchman, dado que no centra tanto su atención a elementos de diseño riguroso, sino en atender a las necesidades e intereses de los *stakeholders*<sup>3</sup> para así tener claridad de los elementos del programa y poder proveer información que sea de utilidad para ellos.

Bajo la perspectiva de Stake, un evaluador no debe juzgar, sino centrarse en diseñar estrategias metodológicas que le permitan recopilar los juicios de las personas relacionadas con el programa

---

<sup>3</sup> El término *stakeholder* es comúnmente utilizado en el argot de la evaluación educativa para denominar a todas aquellas audiencias relacionadas o interesadas con el evaluando, como los beneficiarios, sus administradores, entre otros.

educativo con el fin de proporcionar información que facilite la toma de decisiones para mejorar la manera en que se desarrolla el programa o el evaluando; de esta forma, el evaluador debe reservar sus juicios personales sobre lo que evalúa, ello compete a los especialistas con la información que se ha recopilado. Respecto a la recopilación de datos, no solo observando los efectos buscados (que normalmente son explícitos en la formulación de programas educativos), sino también los efectos secundarios y logros accidentales que puedan presentarse. Esta perspectiva amplía aún más los postulados de los modelos previamente mencionados, ya que más que considerar a la evaluación como un elemento únicamente técnico cuyo interés predominante sea el diseño metodológico, o bien, centrado en todos los posibles elementos que conforman un programa educativo, la evaluación debe ser un proceso de diálogo con los *stakeholders* con el fin de identificar los elementos de interés para la evaluación (de acuerdo con las necesidades de los administradores del programa, de sus usuarios y beneficiarios), centrándose, en particular, en lo que Stake identifica como *la base filosófica de su existir*. Es este elemento el que debe determinar la línea de evaluación, ya que su comprensión por parte del evaluador facilitará la planeación del diseño de evaluación, así como la intención comunicativa de los resultados, procurando que éstos sean acordes con los intereses de los diversos *stakeholders*, ya que son las audiencias quienes comprenden –bajo diversas perspectivas– la filosofía del existir del evaluando.

Comparando este modelo con otras formas de trabajo en evaluación, Stake afirma que la constante búsqueda por la exactitud de las calificaciones es un “recurso cuestionable” (Stufflebeam y Shinkfield, 1993, pág. 241) para emitir juicios que involucren toda la complejidad de un evaluando<sup>4</sup>. Esta crítica se orienta hacia los modelos que centran su atención en la definición lo más exacta posible del diseño metodológico como objetivo principal de una evaluación. Bajo la perspectiva comprensiva, el diseño metodológico orientado a la precisión de medida es relevante, pero no trasciende al punto central enfocado a la comprensión a profundidad de las características del evaluando, en particular respecto a su filosofía y a la base lógica que subyace en él. Para Stake, la evaluación es la constelación de valores entre el complejo de expectativas y criterios que diferentes personas tienen sobre el programa en cuestión.

---

<sup>4</sup> Si bien en este trabajo se hace uso del término *evaluando*, Stake hace referencia a los programas educativos como centro de su propuesta.

A los modelos de evaluación que tienen una fuerte carga teórica y metodológica centrada en la búsqueda de la objetividad, Stake los clasifica como *Evaluación basada en estándares* o *Evaluaciones preordenadas* (Stake, 2006). Como se ha mencionado, este tipo de evaluaciones tienden su atención en la búsqueda de la medición de logros<sup>5</sup>, frente a ello, ser explícito en los objetivos esperados es el primer paso para lograr una medición adecuada, y, por ende, de una evaluación válida y objetiva. Esta situación se relaciona con el énfasis en la racionalidad de la evaluación de este tipo, así como la búsqueda de disminución de sesgos (por parte del evaluador y su perspectiva personal).

Según Stake (2006), hay siete factores que engloban a una evaluación basada en estándares y que definen su forma de conducción frente a un evaluando, a saber: 1. *Receptores*. Implica describir a los beneficiarios, ya que ello determinará los objetivos de la evaluación y las diferentes formas de comunicación de resultados. 2. *Objetivos del programa*. Normalmente entendidos como aquellos que han sido expuestos de forma explícita. 3. *Criterios de evaluación*. Un criterio es un atributo de un objeto o de una actividad que permite conocer sus méritos o deficiencias. Existe una tendencia marcada respecto a que el uso de criterios de evaluación es la única forma de realizar una evaluación válida, ello exige la determinación explícita y objetiva de los mismos. 4. *Estándares de evaluación*. Es la cantidad o nivel de determinado criterio. Es una puntuación, un corte, son los elementos que definen aprobación o suspensión. En ocasiones pueden ser implícitos, ello sucede cuando son rendimientos recomendables (normalmente bajo un juicio de evaluadores o expertos). 5. *Pesos o ponderaciones de síntesis*. Es “la síntesis del proceso por el que se reúne toda la información con el fin de deducir una valoración resumida de la calidad” (Stake, 2006, pág. 111). 6. *Rendimientos o actuaciones del personal y de los participantes*. Son estos elementos los comparados con los estándares. 7. *Costos del evaluando*. La evaluación basada en estándares considera los costos como determinantes para el juicio de valía de un evaluando, dado su orientación centrada en la rendición de cuentas.

---

<sup>5</sup> La primer referencia de Stake respecto a la *evaluación basada en estándares* o *preordenadas* se centra en la búsqueda de medición de logros, sin embargo, no por ello representa que su crítica se enfoque únicamente en el modelo tyleriano. El autor engloba en este tipo de evaluaciones a los modelos que trabajan bajo una lógica estricta de diseño, como el modelo de Suchman.

Generalmente los evaluadores esperan que estos elementos se mantengan inamovibles durante el estudio, con fines de mantener objetividad y una planeación intacta. A pesar las posibles críticas que esta perspectiva pueda presentar, la obtención de mejores datos sobre calidad de un evaluando son aquellos obtenidos por medio de las mediciones del rendimiento, aunque la responsabilidad de un evaluador respecto a la identificación del valor de un evaluando no se alcanzará de mejor forma solo por considerar a un gran número de personas.

Es por lo expuesto previamente que Stake valora la necesidad de incluir una metodología interpretativa a los procesos de evaluación educativa, una metodología más orientada hacia la comprensión de las características del evaluando. En este tipo de evaluación debe considerarse que el evaluando es único, por lo que no puede ser generalizado. Una evaluación comprensiva significa involucrar dentro de la evaluación una interpretación personal, vivir la experiencia de estar *ahí* con el evaluando; por ende, la medición estadística tiene menos peso para la asignación de valor del evaluando, pero ello no significa no sea de utilidad, lo importante es no considerarla como elemento definitorio del valor buscado. Bajo esta perspectiva, la comprensión de un programa se entrelaza con el juicio que se hace de éste, dejando a un lado la objetividad buscada como único referente de verdad por la mayoría de los modelos de evaluación educativa; implica enfrentarse a cuestiones y prácticas humanas, por lo que se vuelve necesario entenderlas como acciones contextualizadas y reconocer que se trabaja con experiencias vividas por personas en un momento y bajo condiciones en particular.

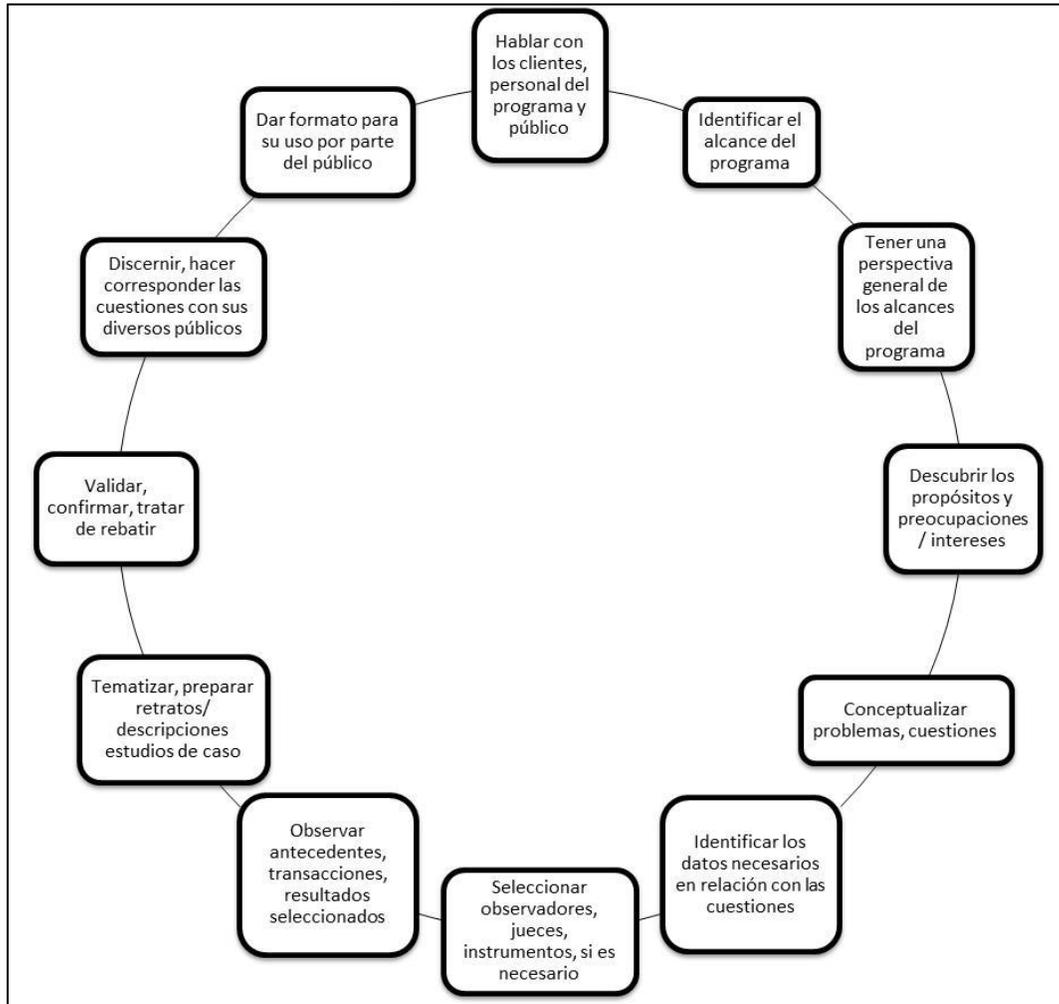
Metodológicamente hablando, es un error pensar que el método de la evaluación comprensiva se basa en un diseño naturalista (puede serlo, pero no necesariamente). Este tipo de evaluación es participe de los llamados métodos mixtos, aunque el uso de pruebas estandarizadas no es frecuente, dado que suele ser insensible a elementos particulares de los evaluados. Bajo esta lógica, este tipo de pruebas pueden ser utilizadas, pero se les da un lugar subordinado, ya que solo se les podrá dar un lugar preponderante cuando realmente puedan informar a la población respecto a la calidad de un evaluando (ya sea que la evaluación se centre en personas, o bien, en un programa o sistema educativo que utilice el rendimiento de los alumnos como indicador de logro). La evaluación preordenada o basada en estándares generalmente sigue un modelo investigativo, centrado en la delimitación de contenidos por observar y de variables por medir

como paso previo a la implementación de la evaluación; mientras que la comprensiva funciona a la inversa, primero se observa de forma natural cómo funciona el evaluando, posteriormente interviene con él para poder hacer juicios sobre su valor (refleja el modelo *estímulo-respuesta* a la inversa).

De manera particular, respecto a las técnicas, la evaluación preordenada prepondera la implementación de hipótesis como guía de trabajo, el uso de los objetivos de comportamiento como guía de diseño y medida y muestreos al azar; mientras que la comprensiva prefiere la observación y el uso de estudios de casos. La forma de comunicación también difiere; mientras la comprensiva es flexible (dado que los acuerdos y estipulaciones previas no son prioritarios), la preordenada es más rígida, siguiendo acuerdos previos. Finalmente, respecto a la valoración, la evaluación preordenada remite a objetivos preestablecidos, mientras que la comprensiva lo incluye, pero no da un juicio “bueno” o “malo”, sino que intenta reflejar todos los puntos de vista sobre el evaluando.

Para tener mayor claridad del modelo comprensivo, la Figura 1.1 muestra “El reloj comprensivo” propuesto por Stake (2006), el cual contiene los acontecimientos que bajo este modelo se deben de considerar al momento del desarrollo de una evaluación. Su lectura no va en el sentido de las manecillas del reloj, sino que únicamente presenta los 12 elementos o pasos que se deben considerar en una evaluación, sin tener un orden establecido de los momentos en los cuales atender uno u otro de ellos, ya que ello dependerá los evaluando y de su contexto en particular.

Figura 1.1 El reloj comprensivo



Fuente: Stake (2006).

Como se puede observar, los elementos metodológicos que enmarcan la evaluación comprensiva no delimitan de forma determinista el diseño que debe guardar una evaluación, ni mucho menos los instrumentos de los cuales debe hacer uso un evaluador para recabar información. La metodología se centra básicamente en implementar algunos de los pasos descritos en el momento que la evaluación lo requiera y a través de los instrumentos que el campo, el presupuesto y el contexto permitan. Derivado de ello, la evaluación comprensiva se centra más en la identificación de los elementos lógicos que subyacen en el evaluando, considerando las opiniones de los involucrados, los intereses de los *stakeholders* y la importancia de la comunicación en todo el proceso de la evaluación.

La evaluación comprensiva es, en esencia, un modelo que recupera elementos valiosos de diversas metodologías en investigación y evaluación educativa, el cual centra su atención en la comprensión del evaluando, más que en la definición metodológica de la búsqueda de objetividad para la emisión de un juicio de valor. Procura recuperar la perspectiva de los *stakeholders* y la filosofía de la existencia del evaluando, con el fin de emitir juicios de valor más centrados en las particularidades de lo observado y en menor medida en la búsqueda de generalización y sistematización que caracteriza a las evaluaciones preordenadas. Bajo esta lógica, es un modelo que si bien fue pensado para la evaluación de programas educativos, el reto consiste en valorar la posibilidad de diseñar evaluaciones de otros ámbitos educativos bajo una perspectiva centrada mayormente en la comprensión del evaluando.

### Reflexiones sobre los usos de los modelos de evaluación

A lo largo del capítulo se han comentado los enfoques, objetivos y diversidad metodológica de los distintos modelos de evaluación. La diversidad de informes existentes en la red respecto a la evaluación de programas educativos y sociales reflejan la multiplicidad de su uso de forma adecuada (por ejemplo, los informes de ADB, 2005; University of Strathclyde, 2004; Western Michigan University, 2000; NRF 2001, entre otros). Sin embargo, también existen las prácticas de evaluación que no necesariamente se desarrollan con una claridad respecto a un enfoque en particular. Citar ejemplos que trabajan de esta manera es complicado, ya que no todos los trabajos de evaluación se publican debido a razones de confidencialidad (como el caso de los CIEES y las evaluaciones de los organismos acreditadores del Consejo para la Acreditación de la Educación Superior, ambos en México). O bien, no hacen explícito su modelo de evaluación, como el caso del Ceneval, cuyas publicaciones técnicas publicadas en su portal (<http://www.ceneval.edu.mx>) no citan modelo de evaluación alguno.

Adicional a la problemática de que no es posible enjuiciar mediante publicaciones los usos adecuados de los modelos de evaluación, se añade la dificultad de la forma en que los usuarios hacen uso de los resultados de evaluación. Como se ha comentado, independientemente de los métodos utilizados, la evaluación tiende en la teoría a realizar esfuerzos en pro de la retroalimentación a la práctica educativa como pilar de su elaboración, sin embargo, las prácticas cotidianas pueden no centrarse en ese fin. Ejemplo de ello es la proliferación del uso de evaluación

de programas con el objetivo único de certificación que favorecen la adquisición de recursos, o bien, el uso de pruebas de ingreso que permiten la distribución de los aspirantes entre los espacios disponibles en las escuelas; ejemplo de ello es el Examen Nacional de Ingreso (EXANI) del Ceneval, prueba que en el año 2014 presentaron 460,752 estudiantes en todo el país con fines de ingreso a la educación media superior (Ceneval, 2015a). Adicionalmente, otro de los ejercicios nacionales en los que la evaluación se usa con fines de certificación del aprendizaje son las pruebas de egreso de nivel licenciatura, las cuales se utilizan para la mejora de indicadores de egreso y titulación de las IES. El uso de estas evaluaciones por parte de las IES no son publicados de manera formal, el menos en México, pero representan el mayor uso de los resultados de las evaluaciones, el cual puede manifestarse mediante los discursos institucionales que refieren a los resultados de las evaluaciones como elemento de reconocimiento cuando éstos son positivos, tal y como se observa en las páginas de Internet de programas de educación superior que han logrado acreditaciones.

Un ejemplo de institución en México que no se enfoca propiamente al uso de la evaluación del aprendizaje de estudiantes solo con fines sumativos es el Instituto Nacional para la Evaluación de la Educación (INEE), el cual en su Plan Maestro de Desarrollo 2007 – 2014 (INEE, 2006) incluye dentro de su definición conceptual de los componentes del sistema educativo una clara influencia del modelo de Stufflebeam, al considerar al contexto, insumos, procesos, productos y objetivos como los elementos que lo definen. Las publicaciones del INEE han procurado abarcar los elementos del sistema educativo bajo la lógica de Stufflebeam, al considerar la evaluación de los *productos*, mediante el uso de pruebas nacionales e internacionales dirigidas a medir y valorar los logros de los estudiantes; al *proceso*, mediante la valoración de prácticas pedagógicas; algunos elementos de *contexto*, observados mediante el análisis de factores asociados al aprendizaje; *insumos*, mediante estudios de infraestructura escolar. Sin embargo, la tendencia de centrarse en los *productos* a través de la observación de los resultados de aprendizaje de los estudiantes ha sido clara; basta con revisar el alto número de publicaciones de esta temática en relación al resto de los elementos que conforman el sistema educativo. Las publicaciones pueden consultarse en: <http://www.inee.edu.mx>.

Esta problemática tiene un impacto dentro del discurso de los modelos de evaluación educativa, ya que, como se ha comentado, dentro del ámbito de la evaluación de los aprendizajes existen

pocas evaluaciones que den cuenta de forma explícita del modelo que define su práctica. El hecho de que no exista una manifestación expresa del modelo que se utiliza (que define la metodología, alcances y límites del trabajo que se realiza), no es un asunto leve, ya que la no definición o explicitación de un modelo de trabajo no permite conocer el enfoque bajo el cual se realizan recomendaciones a los logros alcanzados por los estudiantes, al estado de desarrollo de programas de estudio o los avances de los sistemas educativos, más aún si se considera que los resultados de estas evaluaciones definen aspectos en términos de política educativa, de asignación de presupuesto e incluso en decisiones de vida para quienes las sustentan cuando las evaluaciones son de índole sumativo.

Ahora, si bien los modelos descritos se han desarrollado gracias al interés que los académicos han puesto en pro de la mejora de los procesos de la evaluación de programas educativos, la evaluación a gran escala de los aprendizajes –que también ocasionalmente se usa como elemento de decisión respecto a la eficiencia de programas y sistemas educativos– también debe ser puesta a juicio bajo diversas perspectivas del trabajo evaluativo, ya que históricamente ha seguido un modelo tyleriano que no ha facilitado el progreso de su quehacer, al abandonar elementos valiosos de otros modelos que han sido propuestos a lo largo de los últimos años. Su análisis se vuelve aún más necesario si se considera que ha sido utilizada sin una manifestación explícita del enfoque que la determina.

En ese sentido, el acercamiento realizado a los diversos modelos de evaluación en educación manifiesta la necesidad de considerar un modelo o enfoque que determine la labor de la evaluación del aprendizaje, de forma particular de la evaluación a gran escala, la cual, como se analizará en próximos capítulos, ha enfocado su desarrollo en la mejora metodológica de la medición de rasgos de aprendizaje en los estudiantes, no así en la estructuración de uno o varios modelos que definan su quehacer, y que además, facilita la reproducción de políticas educativas enfocadas en la rendición de cuentas. Adicionalmente, la mención de estos modelos no solo expresa esta necesidad en otro ámbito de evaluación educativa, sino que también permite conocer las diversas maneras de abordar un fenómeno educativo en particular con fines de evaluación. De esta forma, representa la posibilidad de considerar un enfoque existente para

establecer una fundamentación teórica sólida del trabajo de evaluación del aprendizaje a gran escala.

La descripción de los diversos enfoques de evaluación en educación favorece así la posibilidad de la amplitud de un panorama metodológico para las diversas aplicaciones de la evaluación, entre ellas la evaluación a gran escala. Frente a lo expuesto, y con fines de describir de manera más detallada la problemática de enfoque teórico que ha caracterizado la forma en que se ha desarrollado la evaluación a gran escala, en el siguiente apartado se abordan los aspectos básicos de la evaluación a gran escala como una de las formas de la evaluación educativa de mayor auge e impacto para una mejor comprensión de la problemática que subyace a la evaluación a gran escala, la manera en que responde a demandas sociales y políticas, así como su modelo de desarrollo.

## 2. Evaluación a gran escala

La evaluación a gran escala, como una de las formas de evaluación educativa de mayor uso en los sistemas educativos, técnicamente hablando es caracterizada particularmente por su alta afinidad con la psicometría como parte de su metodología; sin embargo, no es su único elemento definitorio, en la práctica cotidiana normalmente sus resultados son utilizados como herramienta de decisión en materia de política educativa y, en algunos casos, como proveedora de información para toma de decisiones que involucran el ingreso o egreso de estudiantes a un nivel educativo en particular. El presente capítulo muestra una descripción de estas vertientes como elementos determinantes de la evaluación a gran escala, asimismo, pretende caracterizar a este tipo de evaluación desde sus inicios y sus tendencias actuales y futuras, detallando los trabajos en México y, en particular, los esfuerzos de evaluación a gran escala del egreso del nivel licenciatura. Esta caracterización permitirá sentar su relación con los modelos de evaluación educativa como antecedente para la delimitación del problema de investigación.

### Antecedentes y conceptualización

La evaluación a gran escala basa su quehacer en la teoría de la medición psicológica (Martínez, 2001), desarrollada desde el siglo XVIII por Laplace y Gauss, aunque su aplicación en educación comenzó en el siglo XIX en países como Alemania, Inglaterra, Estados Unidos y Francia. Es a partir del siglo XX que se desarrolla la *Teoría Clásica de los Tests* (TCT), centrada en elementos estadísticos como confiabilidad y error de medida; esta teoría ha sido uno de los pilares técnicos para el desarrollo de este tipo de evaluación. Los autores principales de la TCT son Spearman y Thorndike (Martínez Arias, 1995, en Martínez, 2001). Es a partir de los años 40 del siglo XX que comienza a desarrollarse fuertemente esta línea de trabajo, prueba de ello es el auge de publicaciones referidas a la psicometría, como la revista *Psychometrika* (1935) y *Educational and Psychological Measurement* (1941). En 1946 Stevens realizó la clasificación de niveles de medición de variables (nominal, ordinal, intervalo y razón), lo cual potenció la implementación de métodos cuantitativos para la medición de rasgos psicológicos en las personas, así como la determinación de medición de variables sociales para el desarrollo de la estadística.

El uso de pruebas psicométricas centradas en la educación inició bajo la intención de medir el aprendizaje como rasgo psicológico, ya que, como se comentó en el Capítulo 1, el aprendizaje era

considerado el elemento que mejor podía reflejar los avances educativos, dado que mostraba de manera directa la influencia de programas sobre los estudiantes. La evaluación a gran escala que comenzó a desarrollarse desde el siglo XX, particularmente en Estados Unidos, tenía como interés identificar si los logros de los estudiantes eran acordes con los mínimos aceptables (Martínez y Santos, 2009). Esto se daba principalmente debido a los procesos tales como la masificación de la educación que comenzó a mediados del siglo XIX. De ahí que la evaluación a gran escala consista en la aplicación estandarizada de pruebas a una gran cantidad de alumnos, ello con la finalidad de identificar el nivel de aprendizaje logrado por un país o región (Martínez, 2008b).

La aplicación estandarizada de pruebas destinadas a medir el logro del aprendizaje bajo el respaldo de técnicas estadísticas ha favorecido la obtención de evidencia empírica objetiva, aspecto de gran aprecio por parte de los *stakeholders*, debido a la confianza que el uso de métodos objetivos trae para la toma de decisiones; por lo que el uso de pruebas de lápiz y papel<sup>6</sup> ha facilitado el alcance de este objetivo, ya que permiten la obtención de información y su eficiente codificación con fines de presentación de resultados. En otras palabras, el uso de técnicas psicométricas y estadísticas ha facilitado el desarrollo de la evaluación, ya que conjuntan un proceso objetivo con una forma eficiente de procesamiento de datos. En particular, la TCT fue de gran importancia para la medición psicológica, pero fue superada por la *Teoría de Generalización* de Cronbach, y por la *Teoría de Respuesta al Ítem* (TRI) de Lazarsfeld y Lord, la cual permite estimar las habilidades de una persona en función de un manejo probabilístico de los datos, a diferencia de la TCT, que se basaba únicamente en respuestas observadas (Martínez, 2001). Actualmente, gracias a estos trabajos ha habido un gran desarrollo técnico respecto a cómo desarrollar este tipo de evaluaciones, ya sea considerando reactivos con respuestas construidas (comúnmente llamados “reactivos de respuesta abierta”), pruebas adaptativas, diseños matriciales<sup>7</sup>, entre otros. Cabe mencionar que el uso de reactivos de respuesta construida no

---

<sup>6</sup> Se hace uso del término “pruebas de lápiz y papel” para referir a pruebas objetivas; sin embargo, en la actualidad existe una tendencia creciente del uso de pruebas objetivas estandarizadas mediante medios electrónicos. El presente trabajo no hace énfasis en esta diferencia, ya que ambas formas de presentación de las pruebas se rigen bajo los mismos estándares, salvo en elementos de aplicación y calificación, donde las pruebas computarizadas permiten acelerar estos procesos, e incluso desarrollar pruebas adaptativas, elemento impensable en el formato lápiz y papel.

<sup>7</sup> Las llamadas *pruebas adaptativas* son instrumentos que, mediante el uso de la TRI, estiman la habilidad de una persona en función de las respuestas que la persona da a un reactivo computarizado. La idea es que la computadora estime la habilidad de la persona para presentarle reactivos cuyo nivel de dificultad sea acorde con su habilidad (Sierra, Valdelamar, Hernández y Sarmiento, 2007). Por otra parte, en el caso de las pruebas *matriciales* son aquellas en las cuales el muestreo de los estudiantes se realiza de forma aleatoria a

implica que la evaluación a gran escala prescinda de su elemento objetivo y sistemático definitorio. Las pruebas que hacen uso de este tipo de formato de reactivos se codifican por medio de rúbricas, cuyo uso para la codificación de respuestas construidas de los estudiantes permite mantener la sistematicidad y objetividad que se presentan en reactivos de formato cerrado.

Si bien ha habido un importante desarrollo técnico en este tipo de pruebas, su uso se ha implementado y mantenido por necesidades educativas y sociales de estandarización de contenidos y su verificación por medio del aprendizaje logrado por los alumnos, alentadas básicamente por decisiones de política gubernamental centrada en la rendición de cuentas, como más adelante se describe. Adicionalmente, la masificación de la educación superior ha traído consigo la necesidad de contar con elementos objetivos para el ingreso a este nivel educativo, y es que las IES requieren de medios de evaluación para asegurar que los estudiantes de recién ingreso cuenten con los conocimientos o habilidades mínimos requeridos para poder enfrentar las demandas que la educación superior implica, además de la imperiosa necesidad de contar con mecanismos que permitan la regulación de los espacios disponibles para el estudio, cuando la solicitud de ingreso alcanza niveles altos cada año<sup>8</sup>. Este proceso se manifiesta de manera similar en el egreso, dado que las IES necesitan medios de verificación de que los estudiantes finalizan su formación profesional con los conocimientos o habilidades mínimas necesarias para poder desarrollarse de manera eficiente en el ámbito profesional, además de contar con medios que favorezcan la regulación del egreso debido a la baja eficiencia de los medios tradicionales, como la tesis o los informes de investigación.

Continuando el análisis de los objetivos de la evaluación a gran escala, Ravela (2006) afirma que las pruebas de este tipo permiten “conocer en qué medida están siendo logrados los conocimientos y competencias que se espera que los estudiantes hayan adquirido cuando finalizan un nivel o ciclo educativo” (pág. 18). Según este autor, existen diversas finalidades de una evaluación a gran escala. La primera de ellas es la acreditación y/o certificación. El término *acreditación* refiere a instituciones y programas, mientras que la *certificación* a las personas. La

---

subconjuntos de reactivos; es decir, solamente algunos estudiantes responden a ciertos reactivos. Al finalizar el estudio, los resultados de todos los estudiantes se conjuntan, infiriendo niveles de habilidad de los estudiantes en todos los reactivos que no hayan contestado (Martínez Arias, 2006).

<sup>8</sup>Por ejemplo, los datos de solicitud de ingreso a la Universidad Nacional Autónoma de México disponibles en: <http://www.escolar.unam.mx> permiten conocer la enorme demanda de este tipo de servicio educativo, al punto que aproximadamente 10% de los aspirantes serán quienes alcancen un lugar para realizar sus estudios profesionales en cada uno de los ciclos en que este proceso se lleva cabo al año.

certificación se centra en la identificación de los conocimientos y competencias estipulados para el nivel educativo en particular. Un segundo objetivo es la *selección*, el cual es utilizado en procesos de ingreso a instituciones educativas o laborales. Otro de sus fines es la toma de decisiones “blandas”, que refiere al sentido formativo de la evaluación. El autor les llama “blandas” debido a que su uso no conlleva una consecuencia formal. En contraparte se encuentra el objetivo de toma de decisiones “duras”, que se relaciona con a las evaluaciones que traen como consecuencia la continuidad de un proyecto o programa. Finalmente, Ravela considera al establecimiento de incentivos y a la rendición de cuentas y responsabilización, que permiten valorar el avance de las acciones educativas de mejora, lo que ha permitido que las evaluaciones estandarizadas hayan tenido un gran auge en el siglo XX.

La clasificación realizada por Ravela refleja a grandes rasgos las diferentes formas de comprender los objetivos de la evaluación a gran escala. Sin embargo, no representa una clasificación rígida con categorías excluyentes entre sí, dado que –de acuerdo con esta categorización– una evaluación cuyos resultados sean condicionante para que los estudiantes puedan egresar del nivel licenciatura o tengan posibilidades de obtener un título profesional puede ser considerada como de certificación, de toma de decisiones “duras”, o incluso, como elemento de rendición de cuentas para efectos de los indicadores que las instituciones utilizan dentro de sus procesos de acreditación.

Por su parte, Martínez y Santos (2009) clasifican a las evaluaciones a gran escala en dos tipos: a) Selección. Se aplican a aspirantes de ingreso a una institución educativa; centran su atención en la validez predictiva respecto al posible desempeño de los aspirantes una vez seleccionados. En su desarrollo privilegia la medición de habilidades, más que de conocimientos e incluye reactivos con dificultad media y, b) Aprovechamiento. Centradas a los alumnos que terminan un ciclo o grado escolar. Este tipo de pruebas se centran en el aseguramiento de la validez y los contenidos con una fuerte carga curricular. Hace uso de reactivos con diferentes grados de dificultad, ya que pretende verificar la diversidad de habilidades de los evaluados. Como se observa, el contexto de la clasificación de Martínez y Santos se establece únicamente bajo una perspectiva centrada en educación básica, lo cual deja fuera otros usos que pudieran ampliar esta clasificación.

Considerando los elementos que se han descrito, los aspectos definatorios de la evaluación a gran escala están supeditados por cuestiones técnicas definidas por la psicometría y usos que los *stakeholders* dan a los resultados. Ahora bien, cabe aclarar una precisión terminológica; en el

idioma inglés a la evaluación a gran escala se le conoce como *large scale assessment*, sin embargo, en dicho idioma existe una diferencia conceptual entre *assessment* y *evaluation*, lo cual permite diferenciar al *assessment* como forma de medición u observación constante de un fenómeno en particular de la *evaluation* como elemento de retroalimentación a partir de las mediciones observadas. En el castellano, se hace uso de término *evaluación* sin diferencia respecto al *assessment*, ya que no existe una traducción fiel para el término. Evidentemente, ello también representa un problema, ya que el término *evaluación* se utiliza de manera indistinta en las prácticas de los evaluadores profesionales al referirse a procesos propios de medición. Sin embargo, más allá del término utilizado para el fenómeno a gran escala en castellano (*large scale assessment vs evaluación a gran escala*), la diferenciación terminológica pareciera no explicar que la base teórico-conceptual de este tipo de estudios se basa en la mejora técnica y no en un modelo que defina su quehacer. Ya que, como la bibliografía muestra, la evaluación a gran escala asume su base psicométrica y su ejercicio en pro de las decisiones que se quieran tomar con los resultados debido a la objetividad que la elaboración de sus instrumentos procura.

### **La evaluación a gran escala como elemento de rendición de cuentas**

Como se comenta en líneas previas, la información generada por las evaluaciones a gran escala es usada como elemento de revisión del avance de sistemas educativos, o bien, con la intención de selección y rendición de cuentas. Este último en particular fue el elemento que dio pauta al uso sistemático de la evaluación a gran escala. De acuerdo con Popham (2001, en Depascale, 2003), la evaluación a gran escala tiene la misión primaria de rendir cuentas, ello a través de la observación de los logros de los estudiantes como elemento de análisis del avance educativo. Según Martínez (2009), la evaluación del aprendizaje a gran escala ha tenido un gran desarrollo gracias a su importancia para el avance de resultados educativos y como insumo para la implementación de políticas de mejora de escuelas.

Actualmente se asume la relevancia social de conocer el avance de resultados educativos con fines de establecimiento de líneas de acción educativas, pero vale recordar que esa perspectiva de la evaluación se ha venido construyendo recientemente. De acuerdo con Anderson (2005), en la segunda mitad del siglo XX algunos gobiernos occidentales comenzaron a enfrentarse a conocer el hecho de que sus estudiantes presentaban un bajo desempeño en evaluaciones académicas. La evaluación entonces dejó de ser un elemento propio de la investigación para convertirse en un

aspecto de atención social. De esta manera, la oportunidad de la educación ya no se entiende de forma estática en lo que respecta únicamente a la preocupación por el acceso a la escuela, sino también se ha vertido hacia una preocupación centrada en que los alumnos alcancen un nivel importante en su formación. De acuerdo con autores como Díaz-Barriga (1994) y Villaseñor (2003), esta preocupación resulta una consecuencia que deriva de las políticas económicas neoliberales que comenzaron en los años ochenta.

Bajo una política neoliberal, la evaluación cumple una función de legalización de la restricción de la educación; es decir, es un reconocimiento administrativo al conocimiento (aunque el instrumento no necesariamente refleje el saber de una persona) (Díaz-Barriga, 1994). La evaluación cumple así entonces una función social relevante, representa una forma de solucionar diferentes problemas, tales como los económicos (en relación a la asignación de presupuesto para los programas e instituciones educativas), de orden social (distribución de personas) y pedagógicos (promover los conocimientos de cada persona). La idea social es la de obtener un conocimiento *objetivo* del saber de cada uno de los estudiantes. Según Calero y Choi (2012), la evaluación a gran escala, que pone énfasis particular en los resultados educativos de los estudiantes, responde fuertemente a una política de rendición de cuentas, en la que se espera conocer el impacto de los *inputs* o insumos que los gobiernos e instituciones destinan a los programas educativos. Esto se da de esta manera debido a que el conocimiento de los resultados de los estudiantes será un indicador que favorezca la toma de decisiones respecto a la inversión que realiza directamente en los sistemas educativos.

Su uso como elemento clave en la rendición de cuentas y en el establecimiento de políticas educativas se ha dado gracias a su metodología centrada en la sistematicidad y la objetividad. Este auge se ha dado en todo el mundo, como se resume en el texto *La calidad de la educación en México: Perspectivas, análisis y evaluación* (SEP, 2002). A continuación se muestran los elementos básicos de evaluación a gran escala en algunos países incluidos en dicho texto.

*Estados Unidos.* Como se ha mencionado, ha implementado políticas de rendición de cuentas que han fomentado el desarrollo de mejoras técnicas para llevar a cabo evaluaciones a gran escala. Actualmente se reconoce la labor del *Educational Testing Service* (ETS) en la implementación de proyectos a gran escala dentro del país, como en proyectos internacionales.

*Holanda.* El Instituto Nacional para la Medición Educativa (CITO, por sus siglas en holandés) elabora y aplica a escala nacional una prueba para la evaluación de todos los alumnos al final de la educación primaria sobre lectura, razonamiento y matemáticas.

*España.* Se creó el Instituto Nacional de Calidad y Evaluación (INCE). El programa de trabajo del INCE incluía la evaluación de la educación primaria y una de la formación profesional.

*Francia.* En este país se aplica el *Baccalaureat*, examen nacional que todos los alumnos que terminan la enseñanza media superior deben aprobar para obtener el diploma correspondiente, con independencia de las calificaciones que hayan obtenido en sus escuelas. Constituye la base para que la autoridad educativa nacional pueda tomar medidas que aseguren un mínimo de calidad en todas las escuelas del país.

*Argentina.* En 1993 se creó un sistema nacional de evaluación, y desde esa fecha se aplican pruebas de alcance y cobertura. Inicialmente a todos los alumnos del último grado de primaria y secundaria, en matemáticas y español, luego se extienden a otros grados e incluyen ciencias naturales y sociales.

*Bolivia.* Estableció un Sistema de Medición de la Calidad (SIMECAL) que comenzó en 1997, aplicando pruebas de rendimiento en los grados 3° y 6° de primaria, en lenguaje y matemáticas.

*Brasil.* En 1993, se estableció el Sistema de Avaliação da Educação Básica (SAEB), el cual no llegó a consolidarse. En 1997 se creó el Instituto Nacional de Estudos e Pesquisas Educacionais (INEP), que evalúa tanto la educación básica y media como la superior. Además de las pruebas para educación básica del SAEB, el INEP aplica el Examen Nacional do Ensino Médio (ENEM) y exámenes de egreso para las principales carreras del primer nivel de la educación superior, el Examen Nacional de Cursos, conocido popularmente como el PROVAO. En el año 2000 el PROVAO cubrió 18 carreras y fue aplicado a casi 198,000 egresados de IES de todo el país.

*Chile.* Las pruebas del Sistema de Medición de la Calidad de la Educación (SIMCE) se aplican a todos los alumnos de 4° y 8° de primaria en español y matemáticas, y a una muestra, en ciencias naturales, geografía e historia.

*Colombia.* El Instituto Colombiano para la Evaluación de la Educación (ICFES) desarrolla pruebas para evaluar el desempeño de muestras nacionales de alumnos de los grados 3° y 5°, y después 7°

y 9° de primaria, en lengua y matemáticas, y luego en ciencias naturales y sociales. Las pruebas SABER y SABER PRO comenzaron a aplicarse en 1991 (las cuales se detallan más adelante).

*Costa Rica.* Colombia ha trabajado con el Instituto de Investigación para el Mejoramiento de la Educación Costarricense (IIMEC) de la Universidad de Costa Rica. Por otra parte, el Ministerio de Educación Pública, con apoyo del Banco Mundial, desarrolló el Centro Nacional de Evaluación para la Educación (CENE-EDU). Ambos centros se encargan del desarrollo de pruebas en los niveles básicos, tanto formativas, como sumativas.

*Ecuador.* El Ministerio de Educación y Cultura estableció un sistema nacional de medición de logros educativos, conocido como APRENDO, cuya primera aplicación de pruebas tuvo lugar en 1996. Las aplicaciones se hacen anualmente, a alumnos de 3°, 7° y 10° grados de educación básica, en lengua y matemáticas. En el año 2012 se creó el Instituto Nacional de Evaluación Educativa (INEVAL), encargado de la evaluación de todo el sistema educativo.

*Guatemala.* En 1997 inició el Programa Nacional de Evaluación del Rendimiento Escolar (PRONERE). Inicialmente se cubrieron las áreas de lectura y matemáticas y se probaron en los grados 3° y 6° de primaria, en escuelas rurales y urbanas, públicas y privadas.

*Honduras.* La Unidad de Medición de la Calidad de la Educación se encarga de las actividades de evaluación. Las evaluaciones cubren los grados 3° y 6° de primaria y las áreas de español y matemáticas.

*Nicaragua, Panamá y El Salvador.* Estos países participan en un proyecto promovido por la Organización de Estados Iberoamericanos (OEI) y la Coordinación Educativa y Cultural Centroamericana (CECC), para el Establecimiento de Estándares para la Educación Primaria en la región, que incluye un componente importante de evaluación.

*Paraguay.* En 1997 se estableció el Sistema Nacional de Evaluación del Proceso Educativo (SNEPE), con la aplicación de pruebas a los alumnos de 6° grado de primaria.

*Perú.* Se estableció la Unidad de Medición de la Calidad (UMC), que comenzó a aplicar pruebas de lengua y matemáticas a los alumnos de 4° grado de primaria en 1996.

*Uruguay.* En 1995 se estableció la Unidad de Medición de Resultados Educativos (UMRE), para encargarse de las actividades de evaluación. Actualmente Uruguay cuenta con el Instituto Nacional de Evaluación Educativa.

Los ejemplos reflejan el auge de la evaluación a gran escala como instrumento de monitoreo del sistema educativo para la implementación de acciones de política educativa. Esta búsqueda de los diferentes sistemas educativos por identificar su avance, particularmente a partir de la observación de los logros de los estudiantes como indicador de ello, implica un propósito particularmente formativo dependiente de la política educativa en cada país. Incluso, aunque los ejemplos mostrados no manifiestan de forma explícita el marco o modelo de evaluación del cual parten, su tendencia respecto a las ideas de Tyler es clara. Por otra parte, las evaluaciones a gran escala mostradas son curriculares, ello debido a que es algo que facilita la retroalimentación de los sistemas educativos, ya que el estándar de evaluación es fijo. De esta forma, como se ha comentado, hacer uso de una prueba curricular bajo el modelo de Tyler facilita la utilización de métodos sistematizados y objetivos que permiten la toma de decisiones políticas y educativas.

Ahora bien, los ejemplos mostrados refieren a evaluaciones dentro de cada país, pero también hay ejercicios importantes de evaluaciones internacionales a gran escala.

En Latinoamérica destacan los trabajos del Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE) de la Oficina Regional de la UNESCO para América Latina y el Caribe (OREALC), centrado en la evaluación de elementos en común de los currículos de los países latinoamericanos participantes en las áreas de Lenguaje (lectura y escritura), Matemáticas y Ciencias; los estudios que se han realizado en el LLECE constituyen el proyecto de evaluación internacional más importante hasta la fecha en Latinoamérica.

Por su parte, la *Association for the Evaluation of Educational Achievement* (IEA) se desarrolló a partir de reuniones realizadas a mediados de los años cincuenta en el Instituto de Educación de la UNESCO, en Hamburgo, ante la insatisfacción por el uso de tasas de graduación o eficiencia terminal como único indicador de la calidad educativa, en 1958 se planteó la posibilidad de un trabajo de evaluación a nivel internacional, que permitiera obtener resultados estrictamente comparables, mediante la utilización de instrumentos equivalentes en contenido y grado de dificultad. Por su parte, la Organización para la Cooperación y Desarrollo Económicos (OCDE) decidió emprender el proyecto denominado *Programme for the International Student Assessment*

(PISA). Las pruebas se aplican cada tres años a estudiantes de 15 años desde el año 2000, centrando la atención en tres grandes áreas: lectura, matemáticas y ciencias, con mayor énfasis en cada una de ellas en las sucesivas aplicaciones<sup>9</sup>. El caso de PISA es especial en lo que a evaluación a gran escala refiere, dado que no es curricular y procura evaluar *literacy*, que en varios países de habla hispana se le ha denominado como *competencia*. La *literacy* no debe ser entendida completamente como *competencia*, ya que refiere al uso que los estudiantes hacen de los aprendizajes logrados en la escuela y que resultan relevantes para desempeñarse adecuadamente en la sociedad actual, a partir de las demandas que ello implica. Evaluar bajo la lógica de PISA permite evitar problemas de acuerdo de contenidos similares entre países con fines de comparación entre sistemas educativos. La intención de la evaluación de PISA trasciende los currículos de los países participantes, y es que es un programa centrado en conocer el avance de los jóvenes para hacer frente a las demandas de las sociedades de la información y del conocimiento. Independientemente del currículo o el modelo del sistema educativo en cada país, se espera que los egresados tengan herramientas que potencialicen su adaptación y éxito en las exigencias actuales (UNESCO, 2005), y PISA ha puesto énfasis en ello. La propuesta de PISA ha tenido un gran impacto en todo los sistemas educativos del mundo y el hecho de que hacer uso de este constructo de medición para sus objetivos le ha permitido no solo liberar esa problemática técnica, sino que ha mostrado la posibilidad de desarrollar evaluaciones a gran escala sin un sentido curricular como base de desarrollo de contenido de medición o de estándares por alcanzar.

Los ejercicios internacionales también son un indicador importante para la observación de los avances de los sistemas educativos, ya que con un estándar en común entre países participantes (ya sea conformado por coincidencias en currículos, como en el caso del LLECE, o bien, por conceptos académicos como estándar, como en el caso de PISA) se pueden realizar tendencias de los resultados de los estudiantes entre ciclos de aplicación, y con ello retroalimentar el impacto de políticas y programas educativos respecto a los avances que se obtengan en países con contextos diferentes y similares al propio.

Si bien estos elementos son una manifestación del enorme avance técnico que este tipo de evaluación ha logrado, vale reflexionar sobre la importancia que los diferentes ministerios de

---

<sup>9</sup> Desde la aplicación de PISA 2012, en algunos países comenzó a evaluarse adicionalmente la Competencia financiera, en la cual México decidió no participar. A partir de la aplicación de PISA 2015 se considerará también en México al área de Solución de problemas en colaboración como elemento de evaluación.

educación otorgan a la evaluación a gran escala del aprendizaje como medida de observación de los avances de los sistemas educativos. Este monitoreo del avance educativo (en particular en educación básica) es una práctica internacional que permite conocer elementos genéricos de la educación para la toma de decisiones en materia de política educativa, dado que permite la identificación de avances o retrocesos en el aprendizaje de la población de un nivel educativo en particular en una dimensión nacional o internacional.

Al respecto, la reflexión que realiza Anderson (2005) permite comprender de mejor manera el interés de los ministerios de educación en la evaluación a gran escala de corte nacional e internacional, al afirmar que en los sistemas de rendición de cuentas basados en resultados reconocen públicamente el desempeño, con el fin de otorgar recompensas (en algunos casos económicas) a las escuelas; de tal suerte que cuando los resultados de las evaluaciones proporcionan mejoras o beneficios para las escuelas o programas educativos, las evaluaciones pueden ser más utilizadas, incluso se pueden considerar como parte integral y necesaria del proceso de toma de decisiones en materia educativa. En el mismo sentido, cabe analizar lo dicho por Hamilton, Stecher y Klein (2002, en Martínez, 2003; pág. 12):

“Los sistemas de rendición de cuentas basados en pruebas se basan en la creencia de que la educación pública puede mejorar gracias a una estrategia sencilla: haga que todos los alumnos presenten pruebas estandarizadas, y asocie consecuencias fuertes a las pruebas, en la forma de premios cuando los resultados suben y sanciones cuando no ocurra así.”

De esta manera, la evaluación (y en particular la evaluación a gran escala) representa una herramienta valiosa para la consolidación de sistemas políticos y económicos basados en la rendición de cuentas. Por supuesto, bajo una política gubernamental globalizadora, ello responde a motivaciones de corte internacional; como lo comenta Cepeda (2009) en su tesis doctoral, como consecuencia de la Declaración Mundial sobre Educación para Todos, celebrada en Tailandia en el año de 1990, la UNESCO (a través de la Oficina Regional de Educación para América Latina y el Caribe, OREALC) desarrolló un discurso sobre evaluación, orientado particularmente en la definición de mediciones que permitieran la comparación entre escuelas e informar a la sociedad sobre los resultados de las mediciones realizadas. Igualmente, el Banco Mundial señaló la relevancia de contar con estadísticas para valorar la calidad de la enseñanza y del aprendizaje; de esta forma se fomentaría que los países participaran en evaluaciones internacionales sobre desempeño educativo, así como para la creación de instituciones evaluadoras que ayudarán a

consolidar una política de rendición de cuentas sobre los resultados obtenidos en educación. De la misma manera, el Banco Interamericano de Desarrollo (BID) recomendó en 1996 la reestructuración de los sistemas educativos, argumentando la necesidad social de centrarlos en el desarrollo de competencias entre escuelas y bajo una política de rendición de cuentas, haciendo así un énfasis particular en la evaluación de la educación, orientándola hacia la participación de los países en proyectos internacionales con fines de obtener puntos de referencia para establecer metas nacionales de aprendizaje.

Claramente en los casos del Banco Mundial y del BID, que son organismos que impulsan la apertura económica de los países en vías de desarrollo a los capitales extranjeros y dado que la descentralización y la evaluación fomentan la apertura económica, sus sugerencias hacia los países en desarrollo se orientarán hacia ese destino (Cepeda, 2009). Las políticas educativas entonces no están alejadas de las políticas económicas globalizadoras, por lo que en los discursos oficiales se busca que la enseñanza favorezca a que los países sean competentes dentro de un mercado mundial. Bajo esta mirada, y como se ha revisado ya, en los años noventa se enfatizó el uso de pruebas a gran escala como el mejor indicador de los resultados de los estudiantes orientados a las competencias requeridas para una competitividad económica. A partir de esta visión es que la evaluación a gran escala de los resultados educativos se comenzó a considerar como una valiosa herramienta de avance de las políticas educativas influenciadas por procesos de globalización económica.

En este sentido, la cultura *mundializadora* (Villaseñor, 2003) que tiene como eje la economía de un Estado permea a otras esferas de lo público, ya que es necesario que las instituciones (entre ellas las educativas) orienten sus esfuerzos para la obtención de los mayores beneficios económicos posibles. Al respecto, algunos de los proyectos de evaluación a gran escala (como los ya mencionados PISA y LLECE) se encuadran en la llamada sociedad del conocimiento, la cual, de acuerdo con Villaseñor (2003) se manifiesta también en la *Economía del conocimiento*. Es decir, el conocimiento que se genere se orienta a la producción de bienes y servicios, así como a su comercialización; de esta forma el conocimiento se convierte en un elemento importante para la obtención de riqueza. Entonces, otro elemento de relevancia para la proliferación en el uso de la evaluación a gran escala también se da en este sentido, ya que es la herramienta que permite valorar si el Estado (en particular los sistemas educativos) están cumpliendo con la función de

preparar a los estudiantes para hacer frente a las demandas de competitividad que la economía del conocimiento implica.

Bajo esta lógica, la evaluación del aprendizaje a gran escala es una herramienta importante para las decisiones políticas relacionadas con la educación, ya que permite que el Estado, manifestado en los sistemas educativos y en la sociedad en general, valore la *calidad* de la educación de sus estudiantes. El término de calidad siempre es ambiguo, aunque algunos de los indicadores para su definición son aquellos a los que Márquez (2004) llama *exógenos*, los cuales refieren a los requerimientos sociales que se espera que cumpla el sistema educativo. De acuerdo con Tiana (2012b), la atención sobre la calidad de la educación comenzó a darse en los diferentes sistemas educativos en medida que los países iban alcanzando la meta de “educación para todos” (pág. 113). En el mundo globalizado posterior a la década de los noventa del siglo pasado, la educación comenzó a visualizarse como un instrumento contra la exclusión y las desigualdades, así como también como un medio para “asegurar su excelencia de manera que todos alcancen resultados de aprendizaje reconocidos y medibles, especialmente en competencia lingüística, competencia matemática y habilidades esenciales para la vida” (World Education Forum, 2000, en Tiana, 2012b, pág. 115). Bajo esta premisa, la calidad irá en función de una estrategia educativa basada en la competencia; al mismo tiempo, la revisión de la calidad de los resultados educativos facilitará la implementación de una política pública de la rendición de cuentas, fuertemente relacionada con la asignación de recursos. Adicionalmente, en lo que respecta a la conceptualización y la metodología que utiliza este tipo de evaluación, se puede apreciar en todos los casos mencionados durante este capítulo que la intención de observar el aprendizaje de los estudiantes como reflejo de la *calidad* de los sistemas educativos es la perspectiva que subyace este tipo de estudios. Dicha perspectiva coincide con el modelo centrado en objetivos propuesta por Tyler en los años cincuenta, la cual, como se describió en el capítulo previo, centra su interés en la observación de logros cuantificables como indicador de avance educativo, considerando el logro de los estudiantes como el indicador más fuerte de ello. Evidentemente, es un modelo de evaluación que favoreció la implementación de la política basada en la rendición de cuentas de un Estado, la cual establece como eje principal de financiamiento la observación del cumplimiento de estándares que el sistema educativo determina como *adecuados*.

En los diferentes programas de evaluación a gran escala mencionados, si bien están centrados mayoritariamente en educación básica y media, se puede observar claramente la tendencia

centrada en objetivos para determinar los logros de los sistemas educativos en los cuales este tipo de evaluaciones se desarrollan. Incluso a través de PISA, la cual a pesar de no ser curricular centra sus objetivos bajo el modelo de Tyler, al considerar el aprovechamiento de los alumnos como factor de valoración de un sistema nacional de educación; además se considera como una herramienta valiosa para la implementación de políticas educativas de corte internacional en los diversos países que participan en ella. Por tanto, resulta importante reflexionar sobre la manera de concebir a la evaluación a gran escala y su papel limitado como elemento de retroalimentación de sistemas educativos, en particular en su desarrollo en México.

### **Evaluación a gran escala en México**

En el caso de México, la historia de la evaluación educativa de manera formal no es tan amplia como en otros países, tal es el caso de Estados Unidos. Sin embargo, los aspectos políticos que dieron pauta a su auge no distan mucho de lo ocurrido en otros contextos. Si bien desde la década de los sesenta en nuestro país comenzaron a utilizarse pruebas estandarizadas para la admisión de estudiantes a instituciones como la UNAM, la UIA y el ITESM (Martínez, 2008a), a partir de los años ochenta, a pesar de los problemas económicos por los que atravesaba el país, México se dio a la tarea de enfrentar la masificación de la educación, así como la necesidad de atender las demandas de competitividad (Martínez, Cepeda, Fuentes y Burgos, 1995; Cepeda, 2009). Para explicar con mayor detalle la manera en que en nuestro país la evaluación (y en particular la evaluación a gran escala) comenzó a cobrar relevancia, vale recordar de forma general las políticas educativas de nuestro país en las últimas décadas.

De acuerdo con la investigación desarrollada por Cepeda (2009), las políticas públicas en relación a la educación y que impactan directamente sobre los procesos de evaluación datan de los años setenta. A inicios de esa década ya se destinaba una gran proporción del Producto Interno Bruto del país a la educación, y si bien se habían hecho enormes esfuerzos de incrementar el acceso a la educación, más de 25% de personas no tenían acceso a educación secundaria. Adicionalmente comienza a plantearse la posibilidad de incrementar la educación básica hasta el nivel secundaria y el tema de la preocupación por la calidad aparece dentro del discurso público. En 1971 se crea el Instituto de Investigación Educativa, encargado de conformar estadísticas sobre la cobertura y reprobación en la educación básica, incluyendo también estudios explicativos sobre los problemas presentados. También se crea la Subsecretaría de Planeación Educativa dentro de la SEP; en 1974

aparece la Subdirección de Evaluación y Acreditación, área que comienza a desarrollar pruebas para el ingreso de alumnos a secundaria (aunque se habían iniciado desde 1972). Con la incorporación de Fernando Solana al frente de la SEP en 1977 comienza la consideración de la evaluación de la calidad de la educación.

Ya en los años ochenta, debido a la crisis económica de principio de década en el gobierno de Miguel de la Madrid, el país comienza una importante apertura económica hacia otras naciones. Como consecuencia, vino un debilitamiento de la industria local, un control de la inflación y la privatización de empresas públicas, así como una dependencia a instituciones crediticias (como el Fondo Monetario Internacional o el Departamento del Tesoro de Estados Unidos). El impacto de esta política en términos educativos se manifestó en el interés notable hacia la calidad, eficiencia terminal y cobertura que el sistema educativo podía lograr. Desde ese momento comienza a valorarse la necesidad de un Sistema Nacional de Evaluación Educativa para sustentar de forma objetiva la toma de decisiones en materia de política educativa enfocada a la calidad. Como consecuencia de esta intención, en el año 1984 se desarrolló el Instrumento de Diagnóstico y Clasificación para el Ingreso a la Educación Normal (IDCIEN).

Ya para el inicio de los años noventa, en el sexenio de Carlos Salinas de Gortari, se formalizó aún más el ingreso de empresas extranjeras a México a través del Tratado de Libre Comercio, además el país se incorporó como nación miembro a la OCDE. Con la finalidad de que el país se mantuviera competitivo frente a otros países en el marco de esos compromisos, se iniciaron varios programas que consideraban a la evaluación como un elemento fundamental en su quehacer. Por ejemplo, se inició el Programa para la Modernización Educativa en 1992, enfocado en una reestructuración del sistema educativo y, a su vez, explicitando la necesidad de identificar las causas de la deserción educativa y el rezago escolar. En esta etapa la Dirección General de Evaluación, Incorporación y Revalidación de la SEP tiene a su cargo la creación del Sistema Nacional de Evaluación Educativa. En este marco se desarrolla el Diagnóstico para Alumnos de Nuevo Ingreso a la Educación Secundaria (IDANIS) –que en este caso es una acción propiamente establecida en la evaluación a gran escala– y el programa de Carrera Magisterial, el cual es un programa de estímulos para docentes (Horbath y Gracia, 2014). También se desarrolla el Programa para Abatir el Rezago Educativo (PARE; cuya duración fue de 1991 a 1995), realizando evaluaciones de aprendizajes en el aula, así como las Olimpiadas del Conocimiento Infantil, que comenzó en 1993, que implicaba ya una evaluación basada en el currículum. Entre los años 1995 y 2000 se realizó el Acuerdo Nacional

para la Modernización de la Educación Básica, así como el Programa Evaluación de la Educación Primaria en 1996, con la idea de generar estadísticas de tendencias educativas de 3° y 6° de primaria. De gran importancia fue la creación de la Ley General de Educación de 1993, la cual, bajo este contexto de competencia económica internacional, resalta a la evaluación de la educación. En el Plan Nacional de Desarrollo de la gestión de Carlos Salinas de Gortari se señaló que el camino por seguir para determinar el financiamiento de las IES sería a través de la evaluación; como consecuencia se instauraron diversos organismos dedicados a la evaluación (por ejemplo, la Comisión Nacional de Evaluación –CONAEVA-, los CIEES, y Ceneval) (Villaseñor, 2003). Dentro de los mayores logros de esta política pública fue la de la coordinación aparentemente lejana del Estado sobre el curso de acción de la educación superior, así como la implementación de un imaginario social de relación entre evaluación y financiamiento; que posteriormente permitiría la asimilación de la sociedad de la evaluación como elemento fundamental de orientación hacia la mejora educativa.

Ya a partir del año 2000, el discurso de cobertura y equidad sigue permeando en las políticas públicas, aunque también se incluye el de integración y funcionamiento del sistema educativo. En particular, en lo que refiere a evaluación, se centró en su consideración como un instrumento de gestión, la separación entre evaluación e investigación, así como la inexistencia de un sistema de información suficiente y adecuado para atender las demandas que la evaluación de la educación tendría (Cepeda, 2009).

Entre los años noventa y dos mil comenzó un especial interés en el desarrollo de instancias externas de evaluación de los resultados educativos de los estudiantes mexicanos. A mediados de los noventa se creó el Ceneval y se unificaron los procesos de selección de alumnos a la educación media superior de la zona metropolitana de la Ciudad de México a través de la Comisión Metropolitana de Instituciones Públicas de Educación Media Superior (COMIPEMS). Bajo la misma lógica, en el año 2002 se crea el INEE, el cual comenzó a desarrollar las actividades de evaluación que la SEP había llevado a cabo desde años antes (Horbath y Gracia, 2014).

En lo que respecta al ámbito de la educación superior se observó la misma tendencia, en general en todo el SES de nuestro país comenzó el desarrollo de diversos programas de evaluación que permitirían reafirmar la política de otorgamiento de recursos públicos. Ya desde el año 2002 Rodríguez (pág. 143) lo comentaba de la siguiente forma: “Todo parece indicar...que la implantación de fórmulas y procedimientos de acreditación de los programas académicos de las

instituciones será una línea de acción prioritaria en la política de educación superior en los próximos años”. Ello se reflejó en una enorme diversidad de programas, tales como los siguientes. *Fondo para la modernización de la educación superior (FOMES)*. Aporta recursos no regularizables para la adquisición y renovación de equipamiento. *Programa Nacional de Superación del Personal Académico (SUPERA)*. Se centra en el otorgamiento de becas para que los docentes realicen estudios de posgrado. *Programa de mejoramiento del profesorado (PROMEP)*. Impulsa la consolidación de cuerpos académicos mediante becas de posgrado y apoyo para profesores de carrera, así como generación de nuevas plazas. *Programa de estímulos al personal docente*. Pretende estimular la calidad, dedicación y permanencia del personal académico mediante estímulos económicos sumados al salario. *Programa de apoyo al desarrollo universitario (PROADU)*. Opera a través de financiamiento a proyectos de colaboración, coloquios, publicación de libros, proyectos de difusión cultural, etc. *Fondos de infraestructura*. El gobierno federal distribuye recursos extraordinarios para construcción y equipamiento de instituciones. *Programas de apoyo a través de CONACYT*. Los recursos que se distribuyen de esta manera incluyen al Sistema Nacional de Investigadores y al Padrón de Posgrados de Excelencia. Este tipo de programas mantienen la perspectiva que enfatiza la combinación de subsidio público con financiamiento centrada en resultados. Incluso, en el marco de la XV Reunión Ordinaria del Consejo de Universidades Públicas e Instituciones Afines (CUPIA) en el año 2000 se aprobó el modelo de financiamiento de las universidades públicas, parte de su relevancia radica en que el elemento central para ello es el desempeño de los estudiantes (Rodríguez, 2002).

Como se puede observar, es a partir de la política educativa a principios de los ochenta que comienza el interés en evaluación como elemento al servicio del Estado, ya que permitiría identificar problemas y jerarquizar las necesidades específicas de los ámbitos educativos, de forma muy similar a como sucedía en otros países. Adicionalmente, desde estos años se vislumbraba ya la creciente demanda de servicios educativos como forma de movilidad social, entre otros beneficios. Frente a estas demandas sociales, bajo la mirada del gobierno, las IES no eran del todo funcionales para los retos globalizantes. Las universidades eran autocomplacientes, permitían la simulación, no se conocían los resultados de sus procesos educativos ni cómo es que éstas invertían sus recursos (Casillas y De Garay, 1993); por lo tanto, como se ha mencionado, se diseñaron estrategias de financiamiento selectivo y condicionado, mediante las cuales se apoyan los proyectos que para el Estado resultan de importancia.

La evaluación tiene entonces un papel crucial en el desarrollo de la política educativa mexicana, dado que se comienza a dar un lugar importante por su función de acreditación de los diversos programas educativos que conforman el SES, además de que se convierte en un elemento de relevancia en consecuencia de los procesos de masificación de la educación. Si se considera la enorme diversidad de currículums profesionales, así como las limitaciones financieras del SES mexicano, la evaluación sentaría entonces un medio de regulación de financiamiento público y de administración educativa que permitiría que los programas educativos se orientaran hacia la calidad. La evaluación educativa, cuando se piensa particularmente como un medio de *acreditación de la calidad* (que, como se ha visto, esa función se ha resaltado en México) centra especial atención al aprendizaje, concebido como la determinación de la “relevancia y pertinencia como la medida en la que se han alcanzado los objetivos de aprendizaje mediante la enseñanza” (Martínez, Cepeda, Fuentes y Burgos, 1995, pág. 202). De esta forma, la evaluación del aprendizaje implicaba la verificación de que el aprendizaje que se pretendía enseñar había sido adquirido por los estudiantes, lo cual constituiría el elemento más relevante para valorar los avances esperados por el SES. Bajo esta perspectiva, el aprendizaje implica entonces la respuesta a problemas concretos y un saber significativo en la realidad social. En ese sentido, se confirma la tesis de Cepeda (2009), al afirmar que las políticas de evaluación forman parte de las políticas educativas, las cuales, a su vez, derivan de las políticas sociales orientadas a atender elementos de interés poblacional; en el caso de nuestro país, la evaluación a gran escala enfocada a los resultados de los estudiantes se convirtió en un valioso instrumento de instrumentación de la cultura de la competencia enfocada a cumplir estándares internacionales. Como consecuencia, autores como De la Cruz, Leyva y Seda (2013), consideran que en la actualidad la evaluación a gran escala es ya una necesidad latente en México como elemento de mejora para las instituciones educativas.

Como se comentaba anteriormente, entre los años 2001 y 2010 nuestro país se caracterizó por un especial interés en la evaluación del aprendizaje a gran escala. Desde 2004 el INEE diseñó los Exámenes de Calidad y Logro Educativo (EXCALE) con el fin de informar sobre el avance del sistema educativo básico como conjunto, no respecto a cada estudiante en lo particular, así como identificar factores asociados al aprendizaje para brindar información para la toma de decisiones políticas. Dado que era una prueba dirigida a conocer el estado general del sistema educativo por medio de la identificación del logro de alumnos, hace uso de diseños matriciales y análisis estadísticos inferenciales para la emisión de resultados.

Por su parte, la Secretaría de Educación Pública (SEP) en el sexenio 2000 – 2006 desarrolló la Evaluación Nacional de Logro Académico en Centros Escolares (ENLACE). La prueba se elaboró con un interés formativo de identificar el logro académico por alumnos de primaria y secundaria en algunas asignaturas clave con base en los planes y programas de la SEP, con el fin de atender las necesidades de cada alumno. Esta evaluación se consideró de forma censal con la finalidad de que los padres de familia y docentes tuvieran información directa sobre los logros alcanzados por los estudiantes, aprovechando así la experiencia que la Dirección General de Evaluación de Políticas de la SEP había alcanzado con evaluaciones desarrolladas desde los años setenta (SEP, 2007). Los objetivos primordiales de este proyecto se centraban no solo en que los docentes y padres de familia tuvieran información de primera mano sobre el avance de los logros educativos de los niños y niñas, sino que esto se desarrollaba con la intención de contribuir a la mejora educativa desde el salón de clases, facilitando con esta información la toma de decisiones del proceso educativo, y, con ello, “fortalecer la convergencia de esfuerzos nacionales y estatales a través del Sistema Nacional de Pruebas de Aprendizaje, de la Política Nacional de Evaluación y del Sistema Nacional de Evaluación” (SEP, 2007; pág. 13). ENLACE se aplicó con reactivos en formato de opción múltiple durante siete años seguidos en todo México a estudiantes tercero, cuarto, quinto y sexto de primaria en las asignaturas: Español, Matemáticas y Formación cívica y ética. También se aplicó en los tres años de secundaria en las mismas asignaturas y cada año de aplicación se centraba en un área en particular.

La prueba ENLACE estuvo siempre sujeta a diversos comentarios por parte de la opinión pública, desde positivos, considerándolo una medición relevante del estatus del logro de los estudiantes, así como negativos, debido a su uso enfocado únicamente como factor de reconocimiento institucional de las escuelas participantes. Adicionalmente se presentaron algunos problemas, desde el factor copia, la dificultad de su aplicación censal y la falta de evaluación de elementos importantes del currículo que no se ajustaran al formato de opción múltiple. Adicionalmente, a pesar de que fue una prueba pensada con fines formativos, su uso se centró en valorar el desempeño de una escuela o de los docentes, aun cuando no es un indicador completo para dicha valoración (Martínez y Santos, 2009). En enero del año 2014, el Secretario de Educación en turno, Emilio Chuayffett Chemor anunció la cancelación de ENLACE, en virtud de que el INEE adquirió autonomía constitucional con la Reforma Educativa del año 2013, promovida por el Presidente Enrique Peña Nieto; de tal suerte que este instituto sería el encargado de desarrollar y aplicar las pruebas educativas en México; frente a ello, en el año 2015 el INEE sustituyó EXCALE y ENLACE

con el Plan Nacional para la Evaluación de los Aprendizajes (Planea). Cabe resaltar que dentro de la Reforma Educativa del 2013, la evaluación de la educación tomó aún mayor fuerza que la mostrada en décadas previas, debido a que se considera que facilita la organización del todo el sistema educativo, así como el conocimiento del estatus *real* de la educación mexicana, ello para la implementación de políticas públicas basadas en evidencias.

De manera particular en educación media superior, en México existe también el Ceneval, institución que desde 1994 comenzó, bajo la iniciativa de la Asociación de Universidades e Institutos de Educación Superior (ANUIES), a fungir como un agente externo de evaluación de la educación superior. Desde el año de su creación el Ceneval desarrolla y califica el Examen Nacional de Ingreso a la Educación Media Superior (EXANI-I) a solicitud de la COMIPEMS (Hernández, Márquez y Palomar, 2006). Dicho instrumento permite desde hace más de 20 años tener un referente objetivo para la toma de decisiones respecto al ingreso a educación media superior en México, ya que actualmente se aplica en las instituciones públicas y privadas en todo el país que así lo solicitan. El EXANI también tiene dos versiones más, una centrada en el ingreso a educación superior (EXANI-II) y otra en estudios de posgrado (EXANI-III). Si bien en cada versión del EXANI se miden elementos distintos y a diferente nivel de complejidad, todos ellos se centran en la evaluación de habilidades razonamiento matemático y verbal, además de otras habilidades como uso de software o dominio de elementos básicos de investigación (para el caso del EXANI-III). Para dimensionar el impacto y la importancia del EXANI dentro de la vida institucional, debe considerarse que en el año 2014 el EXANI-I fue presentado por 460,752 estudiantes (Ceneval, 2015a), el EXANI-II por 729,429 estudiantes (Ceneval, 2015b) y el EXANI-III por 29,092 estudiantes (Ceneval, 2015c); ello considerando todo el país.

Para el caso de la educación superior, además del EXANI, el Ceneval desarrolla otro tipo de instrumentos a gran escala, incluso los elabora de acuerdo con las solicitudes propias de las instituciones educativas. Uno de los proyectos más grandes del centro es el Examen General para el Egreso de la Licenciatura (EGEL), el cual se centra en la medición de conocimiento y habilidades mínimos necesarios para el ejercicio profesional. De esta prueba se hablará a detalle más adelante.

Los ejercicios de evaluación a gran escala en México, al igual que los desarrollados en otros países, han tenido un lugar importante en la educación debido al interés social y político respecto al estado del sistema educativo, ello gracias a la demanda de rendición de cuentas que ha estado en

auge en las últimas décadas y, como se puede observar, se centra particularmente en la educación básica, ya sea de forma muestral o censal. Asimismo, otro de los elementos que han fomentado su desarrollo es la demanda alta de servicios educativos, que tal como lo comentan Martínez, Cepeda, Fuentes y Burgos (1995), es generada por la movilidad social que representa la educación formal. Esta alta demanda implicó la implementación de herramientas de selección de ingreso a servicios educativos, lo cual dio cabida a la generación de exámenes de selección, con la finalidad de que las instituciones pudieran determinar a los estudiantes con mejores puntajes para ingreso a educación media superior y superior, bajo el supuesto de validez predictiva que subyace a estas pruebas (Chaín, Cruz, Martínez y Jácome, 2003). De esta forma, la evaluación a gran escala no cumple con un elemento formativo, sino como una forma de solucionar el problema social de la alta demanda de servicios educativos (Díaz Barriga, 1994).

La evaluación a gran escala en México ha tenido como fundamento de consolidación la intención de toma de decisiones, ya sea como elemento de implementación de las políticas educativas, así como la selección de aspirantes a un nivel educativo; en ambos casos, es una herramienta con fines sociales de alto impacto en la vida de muchos estudiantes, en particular cuando las decisiones de tipo sumativo están relacionadas con sus resultados. Respecto a este último punto, este tipo de decisiones se han establecido de manera formal en muchas instituciones de educación media superior y superior, al punto que los resultados de evaluaciones como de EGEL o EXANI son considerados dentro de su legislación; lamentablemente no hay evidencia formal de su uso como elemento formativo para la mejora de programas o del sistema educativo, salvo por los documentos institucionales dentro de cada universidad, aun cuando –como se ha comentado–, la búsqueda de retroalimentación fue el objetivo manifiesto en los órganos oficiales de coordinación del sistema educativo.

### **Evaluación a gran escala del egreso del nivel superior**

Como se ha podido observar, la evaluación a gran escala en México y en el mundo ha centrado la mayoría de sus esfuerzos en el nivel educativo básico. Sin embargo, también se han realizado algunos esfuerzos para desarrollar estudios centrados en nivel superior. Un ejemplo de ellos en México son los mencionados EXANI-II y EXANI-III del Ceneval, que, si bien son elaborados con fines de selección e incluso diagnóstico previo al inicio de estudios de nivel licenciatura y posgrado, y que son ejercicios relevantes y de alto impacto centrados en niveles de formación profesional, son

pruebas utilizadas con fines de estandarizar los criterios por considerar para lograr el ingreso a educación superior frente a una diversidad de programas de educación media, como medio de regulación de una matrícula restringida para poder atender una alta demanda en la edad escolar correspondiente, así como para la determinación de financiamiento público.

Otro tipo de evaluaciones a gran escala a nivel superior son las relacionadas con el egreso, las cuales representan un elemento de alto impacto para las instituciones educativas y los estudiantes que hacen uso de ellas. Con fines de identificar el estado actual de este tipo de evaluaciones a continuación se presenta un panorama general de los trabajos desarrollados en diversos países, entre ellos México.

Según Nusche (2008), en los últimos 20 años la evaluación de los logros de aprendizaje de los egresados de la educación superior en el mundo ha tenido una gran importancia, debido principalmente a la gran cantidad de interesados en los resultados (estudiantes, empleadores y tomadores de decisiones políticas, entre otros). Los tomadores de decisiones políticas y las instituciones mismas han hecho un esfuerzo por desarrollar instrumentos que permitan obtener información relevante sobre el aprendizaje de los estudiantes, aspecto que se ha logrado en pocos países. De los estudios que se han generado resaltan los siguientes. Se omite la inclusión de los EXANI, previamente descritos.

La Tabla 2.1 que se muestra a continuación incluye una comparación de los elementos básicos de 17 pruebas de diversos países centradas en la evaluación de nivel superior a gran escala.

**Tabla 2.1 Comparativo de evaluaciones a gran escala a nivel superior en diferentes países.**

País	Nombre de la evaluación y año de inicio	Elementos evaluados	Población	Formato y frecuencia de aplicación	Uso de resultados de evaluación
Australia	<b>Graduate Skills Assessment</b> 2000	<p><b>Habilidades genéricas:</b> Pensamiento crítico, Solución de problemas, Comunicación escrita</p> <p><b>Conocimientos y habilidades específicos:</b> No se incluyen</p> <p><b>Resultados no cognitivos:</b> Entendimiento interpersonal</p>	<p><b>Población objetivo:</b> Estudiantes que ingresan y egresados de nivel superior.</p> <p><b>Cobertura:</b> Aproximadamente 2,000 estudiantes de 20 IES de diferentes campos de conocimiento.</p>	<p><b>Formato:</b> Preguntas de opción múltiple y tareas escritas</p> <p><b>Frecuencia:</b> Dos veces al año, una al ingreso y otra al egreso</p>	<p><b>Para sustentantes:</b> Pueden incluir el puntaje en la prueba como parte de su currículum vitae.</p> <p><b>IES:</b> Cuando se trata de alumnos de recién ingreso se identifican aquellos con posibles problemas para apoyarlos. Para los graduados los resultados son un criterio para ingreso a cursos posteriores. En general se utilizan para mejorar el currículum.</p> <p><b>Gobierno:</b> Colecciona información de la calidad de las IES.</p> <p><b>Empleadores:</b> El gobierno promueve los resultados a los empleadores y apoya su uso como herramienta de contratación.</p>
Australia	<b>Course Experience Questionnaire</b> 1993	<p><b>Habilidades genéricas:</b> Solución de problemas, Habilidades analíticas y de comunicación escrita.</p> <p><b>Conocimientos y habilidades específicos:</b> No se consideran.</p> <p><b>Resultados no cognitivos:</b> Entendimiento interpersonal.</p>	<p><b>Población objetivo:</b> Alumnos que cumplan criterios de egreso de las IES.</p> <p><b>Cobertura:</b> Las respuestas se envían a todos los recién graduados. La tasa de respuesta oscila entre 60 y 65%.</p>	<p><b>Formato:</b> Cuestionario</p> <p><b>Frecuencia:</b> Anual</p>	<p><b>IES:</b> Análisis de tendencias, evaluación de programas, desarrollo y mejora de currículum.</p> <p><b>Gobierno:</b> Asegurar la calidad y desempeño de las IES.</p>

Australia	<b>Graduate Destination Survey</b>  1972	<b>Resultados ocupacionales:</b> Resultados de empleo aproximadamente 4 meses después de la graduación (disponibilidad de emplearse, sectores de empleo, salarios anuales promedio, estrategias de búsqueda de empleo) y actividades de estudio futuras	<b>Población objetivo:</b> Alumnos que cumplan criterios de egreso de las IES.  <b>Cobertura:</b> Las respuestas se envían a todos los recién graduados. La tasa de respuesta oscila entre 60 y 65%.	<b>Formato:</b> Cuestionario  <b>Frecuencia:</b> Anual	<b>IES:</b> Análisis de tendencias, evaluación de programas, desarrollo y mejora del currículo acorde con las necesidades del mercado.  <b>Gobierno:</b> Asegurar la calidad y desempeño de las IES
Brasil	<b>Examen Nacional de Cursos</b>  1995-2003	<b>Conocimientos y habilidades específicos:</b> considerados esenciales y comunes en todos los currículos de las IES en un área en particular.  Disponible para 26 áreas del conocimiento.	<b>Población objetivo:</b> Egresados.  <b>Cobertura:</b> Todos los egresado con áreas definidas de estudio. La prueba abarca 70% del egreso nacional.	<b>Formato:</b> Información no disponible  <b>Frecuencia:</b> Anual	<b>Para sustentantes:</b> La participación es obligatoria, la obtención de un diploma se condiciona a la participación. Se dan premios a los mejores resultados.  <b>IES:</b> Las calificaciones más altas se usan con fines comerciales (publicidad y comerciales). Los resultados son un motivador para docentes y alumnos. Desde 2001 los resultados se consideran para acreditación y reacreditación de programas.  <b>Público en general:</b> Los resultados se divulgan en la sociedad para informar sobre la calidad de las IES

Brasil	<p><b>Examen Nacional de Desempeño de estudiantes</b></p> <p>2004</p>	<p><b>Conocimientos y habilidades específicos:</b> Considerados esenciales y comunes en todos los currículos de las IES en un área en particular. Disponible para 13 áreas del conocimiento.</p> <p><b>Conocimiento de contenidos generales:</b> Diversidad biológica y social, políticas públicas, redes sociales, ciudadanía y problemas de eventos recientes.</p> <p><b>Habilidades genéricas:</b> Habilidad para inferir, interpretación de textos poéticos, establecimiento de puntos de vista comunes, identificar asociaciones, deducción y entendimiento de gráficos.</p>	<p><b>Población objetivo:</b> Estudiantes que ingresan (con menos del 22% de los créditos aprobados) y egresados (alumnos con más del 80% de créditos aprobados).</p> <p><b>Cobertura:</b> En 2004 la muestra inicial del proyecto incluyó a 140,340 estudiantes. Se consideró a 51% de los estudiantes inscritos y 69% de los egresados</p>	<p><b>Formato:</b> Reactivos objetivos y preguntas de ensayo</p> <p><b>Frecuencia:</b> Una vez cada 3 años</p>	<p><b>Estudiantes:</b> Obligatoria para quienes resulten seleccionados, se condiciona el diploma a su participación. Se dan premios a los mejores resultados.</p> <p><b>IES:</b> Análisis de tendencias, evaluación de programas, desarrollo y mejora de currículo.</p> <p><b>Gobierno:</b> Los resultados es uno de los indicadores considerados en la evaluación de las IES.</p> <p><b>Empleadores:</b> Pueden preguntar acerca de los resultados de una persona como evidencia objetiva del desempeño de una persona en un área de conocimiento.</p>
Canadá	<p><b>National Graduate Survey</b></p> <p>1978, en 1987 surgió adicionalmente una encuesta de seguimiento de egresados</p>	<p><b>Competencias ocupacionales:</b> Resultados de empleo las 2 y 5 años de egreso. Información del número, características y duración de los trabajos desde el egreso. Nivel de satisfacción con la experiencia en las IES.</p>	<p><b>Población objetivo:</b> Egresados de todas las IES públicas a 2 años del egreso; la encuesta de seguimiento se da a los 5 años del egreso.</p> <p><b>Cobertura:</b> Se seleccionan entre 35,000 y 40,000. La tasa de respuesta es de 65%</p>	<p><b>Formato:</b> Entrevista telefónica asistida por computadora.</p> <p><b>Frecuencia:</b> Aproximadamente cada 5 años, con un seguimiento cada 3.</p>	<p><b>IES:</b> Análisis de tendencias, evaluación de programas, desarrollo y mejora del currículo, aunque no se dan resultados de cada IES, sino que se reportan los resultados del empleo de los egresados</p> <p><b>Gobierno:</b> Planeación educativa.</p>

Canadá	<b>Youth in Transition Survey</b>  2000	<b>Competencias ocupacionales:</b> Preguntas sobre el tercer ciclo educativo (jóvenes entre 22 y 24 años de edad) incluye cuestionamiento sobre educación postsecundaria y resultados de empleo	<b>Población objetivo:</b> Dos cohortes: 15 y 18-22 años (no se restringe a estudiantes o a egresados).  <b>Cobertura:</b> Participación voluntaria. La muestra inicial fue de 29,000 personas. La tasa de respuesta de la cohorte 18-20 es de 80%	<b>Formato:</b> Entrevista telefónica asistida por computadora  <b>Frecuencia:</b> Cada dos años	<b>Desarrollo de Recursos Humanos Canadá (HRDC):</b> Recolecta información sobre las influencias de la juventud, transiciones en la vida de los jóvenes. Provee información a educadores y a analistas sociales y políticos
México	<b>Examen General para el Egreso de la Licenciatura</b>  1994	<b>Conocimientos y habilidades específicos:</b> Considerados esenciales y comunes en todos los currículos de las IES en un área en particular.  Disponible para 40 áreas del conocimiento	<b>Población objetivo:</b> Estudiantes que hayan cumplido el 100% de créditos (no egresados de programas técnicos).  <b>Cobertura:</b> Depende del campo de conocimiento, más de 15,000 sustentantes al año.	<b>Formato:</b> Reactivos de opción múltiple  <b>Frecuencia:</b> Los estudiantes o instituciones se registran para cuatro aplicaciones nacionales predeterminadas, aunque cada IES puede solicitar una aplicación en el momento que lo decida.	<b>Estudiantes:</b> Provee información de su desempeño en relación a un estándar nacional. Se dan certificados a estudiantes con alto desempeño. Se entrega una medalla para los mejores resultados cada cuatrimestre.  <b>IES:</b> Análisis de tendencias, evaluación de programas, desarrollo y mejora de currículo. Algunas IES usan los resultados como criterio adicional de titulación.  <b>Empleadores:</b> Pueden preguntar a los candidatos sus resultados como evidencia objetiva del desempeño en un área en particular.
México	<b>Examen General para el Egreso del Técnico Superior Universitario</b>  2000	<b>Conocimientos y habilidades específicos:</b> Niveles de comprensión y habilidades de solución de problemas necesitados en el ámbito profesional. Existen 19 pruebas disponibles para las diferentes formaciones tecnológicas.  <b>Conocimientos y habilidades genéricos:</b> Conocimientos y habilidades necesarias para todas las carreras.	<b>Población objetivo:</b> Estudiantes de programas técnicos que hayan cumplido el 100% de créditos.  <b>Cobertura:</b> Todos los egresados de las universidades tecnológicas de todo el país.	<b>Formato:</b> Reactivos de opción múltiple.  <b>Frecuencia:</b> Los estudiantes o instituciones se registran para aplicaciones nacionales predeterminadas.	<b>Estudiantes:</b> Provee información de su desempeño en relación a un estándar nacional.  <b>IES:</b> Análisis de tendencias, evaluación de programas, desarrollo y mejora de currículo. Algunas IES usan los resultados como criterio adicional de titulación.  <b>Empleadores:</b> Pueden preguntar a los candidatos sus resultados como evidencia objetiva del desempeño en un área en particular.

Reino Unido	<p><b>Destination of Leavers from Higher Education</b></p> <p>2002</p>	<p><b>Competencias ocupacionales:</b> Resultados de empleo y estudios posteriores seis meses después del egreso, los tipos de trabajo en los que se encuentran y cuánto tiempo pasará para tener nuevos estudios</p>	<p><b>Población objetivo:</b> Recién egresados de IES públicas que hayan obtenidos calificaciones relevantes y que hayan estudiado medio tiempo o tiempo completo (6 meses después del egreso).</p> <p><b>Cobertura:</b> Se envían cuestionarios a todos los egresados elegibles. La tasa de respuesta es de 77% en estudiantes de tiempo completo y de 70% en estudiantes de medio tiempo.</p>	<p><b>Formato:</b> Cuestionario</p> <p><b>Frecuencia:</b> Anual</p>	<p><b>IES:</b> Análisis de tendencias, evaluación de programas, desarrollo y mejora de currículo, aunque No se dan resultados de cada IES, sino que se reportan los resultados del empleo de los egresados.</p> <p><b>Gobierno:</b> Evaluación y planeación.</p>
Estados Unidos	<p><b>Collegiate Assessment of Academic Proficiency</b></p> <p>1988</p>	<p><b>Habilidades genéricas:</b> Escritura, lectura, matemáticas, ciencia, razonamiento, pensamiento crítico y contenido curricular transversal a todos los campos de conocimiento.</p>	<p><b>Población objetivo:</b> Todo tipo de estudiantes.</p> <p><b>Cobertura:</b> Entre 1988 y 2001, el test se ha utilizado por más de 600 IES y 450,000 estudiantes.</p>	<p><b>Formato:</b> Reactivos de opción múltiple y preguntas de ensayo</p> <p><b>Frecuencia:</b> Flexible</p>	<p><b>Para sustentantes:</b> Las IES pueden promover incentivos para los sustentantes. Se hacen preguntas de motivación para determinar la seriedad con la que se toma la prueba.</p> <p><b>IES:</b> Análisis de tendencias, demostración de la efectividad de programas y su mejora a través del tiempo, comparación de resultados de estudiantes con un referente nacional.</p> <p><b>Estados:</b> Se publican resultados por estado con fines de comparación.</p>
Estados Unidos	<p><b>Measure of Academic Proficiency and Progress</b></p> <p>1988</p>	<p><b>Habilidades genéricas y específicas:</b> Lectura y pensamiento crítico en contextos de humanidades, ciencias sociales y ciencias naturales. Escritura, matemáticas.</p>	<p><b>Población objetivo:</b> Estudiantes sin distinción de avance.</p> <p><b>Cobertura:</b> Ha sido utilizada por más de 1 millón de sustentantes.</p>	<p><b>Formato:</b> Reactivos de opción múltiple y preguntas de ensayo</p> <p><b>Frecuencia:</b> Flexible</p>	<p><b>Para sustentantes:</b> Las IES pueden promover incentivos para los sustentantes.</p> <p><b>IES:</b> Análisis de tendencias, demostración de la efectividad de programas y su mejora a través del tiempo, comparación de resultados de estudiantes con un referente nacional.</p>

Estados Unidos	<b>Tasks in Critical Thinking</b>  1992 (ya no es vigente)	<b>Habilidades genéricas y específicas:</b> Habilidades de investigación, análisis y comunicación. Desempeño basado en tareas de diversas disciplinas (ciencias sociales, ciencias naturales, humanidades y artes).	<b>Población objetivo:</b> Todo tipo de estudiantes.  <b>Cobertura:</b> Entre 1992 y 2001 la prueba se había aplicado a 35 IES con entre 200 y 500 sustentantes en cada institución.	<b>Formato:</b> Reactivos de respuesta abierta y tareas basadas en la resolución de problemas  <b>Frecuencia:</b> Flexible	<b>Para sustentantes:</b> Las IES pueden promover incentivos para los sustentantes.  <b>IES:</b> Análisis de tendencias, demostración de la efectividad de programas y su mejora a través del tiempo, comparación de resultados de estudiantes con un referente nacional.
Estados Unidos	<b>Major Field Tests</b>  1990	<b>Conocimientos y habilidades específicas:</b> Consideradas las más importantes en cada campo de estudio: conocimiento factual, habilidad para analizar y solucionar problemas, habilidad para entender relaciones, habilidad para interpretar gráficas, diagramas y tablas en cada campo de conocimiento. Disponible para 15 áreas del conocimiento y también para la maestría en administración de negocios.	<b>Población objetivo:</b> Estudiantes de últimos años.  <b>Cobertura:</b> Más de 500 IES utilizan el test cada año, se aplica a más de 70,000 personas	<b>Formato:</b> Reactivos de opción múltiple  <b>Frecuencia:</b> Flexible	<b>Estudiantes:</b> Los resultados se consideran como parte de un curso de los últimos semestres como parte de un requisito de egreso.  <b>IES:</b> Los resultados se utilizan para decisiones de medio impacto. Permite conocer la efectividad del currículo. Análisis de tendencias, demostración de la efectividad de programas y su mejora a través del tiempo, comparación de resultados de estudiantes con un referente nacional.
Estados Unidos	<b>Collegiate Learning Assessment</b>  2002	<b>Habilidades genéricas y específicas:</b> pensamiento crítico, razonamiento analítico, comunicación escrita, habilidad para el uso de la información.  <b>Competencias:</b> Tareas de la vida real, tal y como la preparación de un oficio o el uso de diversos documentos. Desempeño basado en tareas en el contexto de diversas disciplinas (ciencias naturales, ciencias sociales, humanidades y artes).	<b>Población objetivo:</b> Estudiantes de recién ingreso y de los últimos años.  <b>Cobertura:</b> Entre 2002 y 2005 participaron 134 instituciones	<b>Formato:</b> Tareas basadas en desempeño  <b>Frecuencia:</b> Anual	<b>Para sustentantes:</b> Algunas IES dan premios a sus estudiantes con base en los resultados.  <b>IES:</b> Los resultados se utilizan para decisiones de medio impacto. Permite conocer la efectividad del currículo. Análisis de tendencias, demostración de la efectividad de programas y su mejora a través del tiempo, comparación de resultados de estudiantes con un referente nacional.

Estados Unidos y Canadá	<b>National Survey of Student Engagement</b>  2000	<b>Resultados no cognitivos:</b> Información sobre el compromiso estudiantil (cómo los estudiantes invierten su tiempo y lo que obtienen de sus cursos, actividades extracurriculares y servicios de las IES).	<b>Población objetivo:</b> Estudiantes de recién ingreso y de los últimos años.  <b>Cobertura:</b> Han participado más de 1,160,000 estudiantes	<b>Formato:</b> Cuestionario  <b>Frecuencia:</b> Anual	<b>IES:</b> Desarrollo y mejora de currículo. Análisis de tendencias, demostración de la efectividad de programas y su mejora a través del tiempo, comparación de resultados de estudiantes con un referente nacional.  <b>Público en general:</b> Los resultados generales están disponibles para consultar la experiencia de los estudiantes en las IES.  <b>Gobierno:</b> Los datos pueden ser utilizados como indicador de efectividad institucional en procesos de acreditación.
Centros de pruebas en Estados Unidos, Canadá y otros países	<b>Graduate Record Examination</b>  1996	<b>Habilidades genéricas:</b> Razonamiento verbal, cuantitativo y escritura analítica.	<b>Población objetivo:</b> Egresados que pretenden ingresar a escuelas profesionales.  <b>Cobertura:</b> Información no disponible	<b>Formato:</b> Reactivos de opción múltiple y tareas escritas  <b>Frecuencia:</b> Flexible	<b>Para sustentantes:</b> La participación es obligatoria para ingresar a determinados programas educativos.  <b>IES:</b> Comparar el desempeño de los sustentantes para facilitar la decisión de selección de aspirantes.
Colombia	<b>Examen de estado de calidad de la educación superior</b>  2009	<b>Habilidades genéricas:</b> para 24 campos del conocimiento	<b>Población objetivo:</b> Estudiantes que hayan aprobado por lo menos el 75% de los créditos académicos del programa que cursan.  <b>Cobertura:</b> No se especifica.	<b>Formato:</b> Reactivos de opción múltiple  <b>Frecuencia:</b> Información no disponible.	<b>IES:</b> Indicadores y análisis de tendencias y comparación entre programas e IES.

Fuente: Elaboración propia con información de Nusche (2008) y Kuhn, Zlatkin-Troitschanaskaia y Bergsman (2011).

La Tabla 2.1 permite observar la forma en que las pruebas a gran escala son utilizadas, que de acuerdo con los discursos oficiales, se orientan hacia la retroalimentación de las instituciones educativas a nivel superior. En general, la educación superior no es obligatoria en ninguna región del mundo, aunque su relevancia también se atañe a que actualmente la educación durante toda la vida es considerada como un derecho (UNESCO, 2005), además que, como se ha comentado, es un elemento clave en la intención de los Estados basados en la rendición de cuentas y en la economía globalizadora de alcanzar estándares de competitividad específicos; es por ello que en el discurso oficial cobra sentido que los diferentes sistemas educativos busquen la manera de retroalimentar la situación de este nivel de estudios a pesar de que la condición de no obligatoriedad facilite el desarrollo de currículos múltiples, con el uso de diversas concepciones y metodologías pedagógicas para cada una de las disciplinas profesionales. Si bien en la mayoría de los casos suelen haber elementos o temáticas en común, la falta de unificación curricular ha sido un problema importante para la evaluación de este nivel educativo. A pesar de ello, los estudios mostrados reflejan dos tendencias claras para realizar la evaluación del avance de este nivel educativo: la observación de habilidades en un profesionista egresado (ya sean *genéricas*, disciplinares o una combinación de ambas), o bien, bajo una modalidad de encuesta respecto a varios temas, pero particularmente centrada en la empleabilidad. Por supuesto, que los instrumentos se enfoquen primordialmente en la empleabilidad es relevante, dado el interés de los egresados en ello, aunque también refleja la relevancia que los diversos Estados dan a los *resultados* de la formación profesional como reflejo de *calidad*.

Los resultados de las evaluaciones enlistadas centran su atención en la retroalimentación de los sistemas educativos a través de envío de información a los ministerios de educación para ello, así como análisis de tendencias para las instituciones de educación superior. Adicionalmente, un elemento importante es que a este nivel se busca dar resultados a los estudiantes (en particular en aquellas en las que las pruebas no son censales).

Adicional a lo ya comentado, pareciera haber un patrón claro de evaluación en este nivel educativo; si bien hay algunos estudios en los cuales no se hace distinción del avance de los estudiantes (como los casos de la *Youth in Transition Survey* de Canadá, el *Collegiate Assessment of Academic Proficiency* y el *Measure of Academic Proficiency and Progress* y el *Tasks in Critical Thinking* todos de Estados Unidos), la mayoría de los estudios revisados se centran en egresados de estudios técnicos o universitarios, o bien, en estudiantes próximos a egresar.

Los elementos derivados de esta comparación permiten identificar claramente que la tendencia para la retroalimentación de los avances de este nivel educativo mediante una metodología a gran escala tiene grandes coincidencias con este tipo de evaluación en nivel básico y medio superior previamente descrito. En particular, se centran en la medición de los logros de los estudiantes (ya sea en cuestiones disciplinares básicas obtenidas por medio de un análisis de los aspectos comunes entre currículos o en habilidades que pueden desarrollarse independientemente del área de estudio, así como por sus logros respecto a su empleabilidad y estudios posteriores) como elemento que permite valorar el avance o situación de la educación a nivel superior, con la diferencia que aquí los estudiantes pueden obtener incluso premios o valores curriculares adicionales a su formación por su desempeño mostrado<sup>10</sup>. Independientemente de este último aspecto, es claro que la perspectiva que subyace en estas evaluaciones tiene una tendencia basada en la evaluación de objetivos de Tyler, ya que no reflejan elementos de proceso<sup>11</sup>, actitudinales o de otros elementos que pudieran complementar la visión respecto a la calidad de la educación superior. Asimismo, hay una clara tendencia hacia el aseguramiento de la calidad de la formación que se otorga. De esta forma, tal como lo comenta Díaz Barriga (1994), se fomenta el esquema neoliberal de la educación, que, como se ha comentado, pretende hacer equivalente la calidad de la educación a las demandas del Estado con fines de hacerse económicamente más competente.

A manera de cierre vale resaltar que la evaluación a gran escala se ha convertido en una herramienta importante para los sistemas educativos, ya que ha sido insumo para valorar el impacto de políticas educativas, su implementación mediante mecanismos de financiamiento y la definición de lineamientos educativos mediante la identificación de los logros alcanzados por los estudiantes, particularmente en elementos cognitivos o conceptuales. En el caso de la evaluación a gran escala en nivel superior, ésta ha seguido una línea similar en su desarrollo, enfocándose principalmente en los logros obtenidos por los egresados en cuanto a contenidos académicos. Si bien esta herramienta de análisis educativo es eficiente y de gran utilidad gracias a la técnica de gran valía y complejidad psicométrica que ha desarrollado, pedagógicamente hablando no ha reflejado un avance respecto al modelo que le subyace, como sí ha sucedido en otros tipos de evaluaciones, en particular con la evaluación de programas educativos (como se comenta en el

---

<sup>10</sup> Mismo que puede ser considerando como elemento para fomentar la participación de los egresados.

<sup>11</sup> Si bien los casos de Canadá y Estados Unidos pretenden identificar avances en todos sus estudiantes, no solo en los egresados, el tipo de información obtenida refiere también a los logros alcanzados en diferentes momentos, lo cual no difiere sustancialmente de las evaluaciones al momento del egreso.

Capítulo 1), donde se ha podido acrecentar su visión de la evaluación con fines de obtención de información más completa que amplíe las perspectivas del evaluando para su análisis a mayor profundidad y, con ello, dar una retroalimentación considerando varios factores además del logro de estudiantes. Centrándose en la educación superior, la evaluación de su egreso ha sido un elemento desarrollado por varios países como indicador de relevancia para la acreditación de los programas educativos, así como para la búsqueda de orientación de los currículos hacia las características deseables de calidad determinadas por el Estado, ello gracias a que permite una visión a nivel macro que complementa las decisiones institucionales, así como aquellas que reorientan las prácticas didácticas dentro del aula. Sin embargo, el valor de sus resultados como elemento de decisión de retroalimentación de la formación profesional puede ser fácilmente cuestionado y con ello poner en duda su utilidad como indicador fiable de evaluación.

Considerando este panorama internacional y, en particular en México, el siguiente capítulo tiene como finalidad mostrar a más detalle la problemática que ha representado considerar a la evaluación a gran escala del egreso de nivel superior bajo un solo enfoque, las limitaciones que ha tenido su implementación en México, la manera en que se han utilizado sus resultados, y, con ello, mostrar la oportunidad que las diversas posturas de evaluación representan para la mejora de este tipo de evaluación.

### 3. Descripción del problema y metodología

Los capítulos previos han mostrado de forma general las diferentes maneras de abordar los fenómenos educativos con fines de evaluación, así como los ejemplos más representativos de la evaluación a gran escala en diversas partes del mundo, y, en especial en México. Se comentaron ya sus usos como elemento constitutivo de una política basada en la rendición de cuentas, algunos inconvenientes de la evaluación que se desarrolla sin un bagaje claro del modelo que le subyace, así como de algunos posibles usos de la evaluación a gran escala que pueden no favorecer del todo a los *stakeholders* ni a la educación misma. Los usos de la evaluación a gran escala por parte de la sociedad en general y por las IES pueden ser diversos e incluso nocivos, y es que valorar a la evaluación a gran escala únicamente bajo la perspectiva de sus avances en la precisión técnica de medida de constructos psicológicos puede potencializar su utilización solo como un aspecto de medición. Un ejemplo de este tipo de uso es el *ranqueo*, y es que hay una tendencia actual en jerarquizar a las instituciones o programas educativos en función de los resultados de los estudiantes en evaluaciones objetivas. Normalmente este tipo de juicios son descontextualizados de la enseñanza (Martínez, 2009) y no favorecen una mejora institucional, al no dar una retroalimentación de utilidad para las IES, los programas educativos o las mismas instituciones. Por supuesto, el uso del *ranqueo* representa también un uso de la evaluación enfocado únicamente en resultados concretos de una forma simplificada. Con tan solo enlistar los resultados de una prueba no se lleva a cabo un proceso de evaluación, sino una simple jerarquización de los resultados de una medición que, al igual que otros métodos, tiene limitaciones en la observación de un fenómeno en particular. Más aún, resulta peligroso juzgar a todo un programa educativo o institución a partir de los resultados de algunos de sus estudiantes.

El claro ejemplo de esta práctica es la desarrollada por el Ceneval con el *Padrón de programas de licenciatura de alto rendimiento académico*. El padrón es un listado de programas académicos de formación profesional que cumplen con los criterios del Ceneval para considerarse como de alto rendimiento académico, en particular, a partir de los resultados obtenidos en el EGEL. El padrón clasifica en niveles 1 y 2 a los programas educativos cuyos egresados obtienen buenos resultados; si entre 60 y 80% de los estudiantes obtienen buenos resultados, se otorga nivel 2; si al menos 80% de los estudiantes obtienen buenos resultados, el Ceneval les otorgará el nivel 1<sup>12</sup>. Desde el

---

<sup>12</sup> Más información en: [http://idap.ceneval.edu.mx/portal\\_idap/principal.jsf](http://idap.ceneval.edu.mx/portal_idap/principal.jsf)

año 2011 comenzó la implementación de este padrón, el cual integra los programas educativos de nivel universitario (para todos los EGEL disponibles) y anualmente otorga un reconocimiento a los programas cuyos egresados obtienen resultados satisfactorios o sobresalientes<sup>13</sup> en el EGEL que sea acorde con el programa educativo en cuestión. Cabe aclarar que la inscripción a la participación el padrón está condicionada a que el total de los egresados o una muestra estadísticamente representativa de los egresados (seleccionada por el Ceneval) presente el EGEL cada año. Una vez que el programa cumple con esta condición, su inclusión en el padrón dependerá de los resultados obtenidos por los estudiantes. La práctica conlleva algunas problemáticas, específicamente en dos sentidos. El primero de ellos es que debido a que se clasifica a un programa como *de calidad* basándose únicamente en los resultados de una prueba que: a) no es curricular<sup>14</sup>; b) no presentan todos los egresados de todas las instituciones del país, lo que imposibilita una comparación válida entre programas educativos; c) independientemente de su calidad en estructura conceptual y en los reactivos que la conforman, su carácter nacional la orilla a correr el riesgo de descontextualizarse de la realidad de enseñanza y; d) no la presentan todos los egresados del programa en cuestión<sup>15</sup> en todas las IES. El segundo problema refiere a no considerar otros elementos para hacer un juicio de *calidad*, y, más aún, de clasificar en niveles a los programas educativos. En este aspecto el Ceneval perdió de vista que una prueba tiene limitaciones respecto a su espectro de valoración, al punto que para poder hablar de *calidad* educativa no bastará con observar únicamente los resultados que obtengan los estudiantes que hayan participado en ella. De esta forma, la creación de este padrón representa una búsqueda del uso de una evaluación a gran escala con fines de acreditación de programas educativos, el cual es resultado de una búsqueda de implementación del uso de la evaluación con fines de rendición de cuentas de las instituciones, práctica que aporta poco a la finalidad retroalimentadora de la evaluación en educación.

Otro de los problemas que enfrenta la evaluación del egreso a gran escala, en particular la relacionada con el egreso del nivel licenciatura, es la llamada *enseñanza para la prueba*

---

<sup>13</sup>Si bien en el próximo capítulo se hace una descripción detallada de los elementos que conforman al EGEL, baste ahora con mencionar que los resultados de cada sustentante se clasifican en tres niveles de desempeño: Aun no satisfactorio, Satisfactorio y Sobresaliente.

<sup>14</sup>Este aspecto pudiera ser una ventaja en algún momento si se logra la evaluación de habilidades o competencias disciplinares afines entre todos los programas de estudio que hacen uso de la prueba.

<sup>15</sup>Salvo en las instituciones cuya reglamentación obliga a que todos los estudiantes presenten la prueba como requisito de egreso, tal es el caso de la Universidad Autónoma del Estado de Hidalgo o la Universidad Autónoma de Aguascalientes.

(DePascale, 2003), la cual consiste en que los educadores centran su enseñanza en los temas que se incluyen en los reactivos de las pruebas estandarizadas. Una de las razones de ese problema es el hecho de que no exista un currículum unívoco para el nivel educativo, por ende, el contenido de la prueba se vuelve un contenido de facto para la enseñanza, si se suma un énfasis particular en la rendición de cuentas, en el financiamiento o en la intención de ranqueo como objetivo primario, el problema se incrementa. Esta situación está cada vez más en aumento en México, debido a la importancia sumativa o de acreditación que han tenido las evaluaciones a gran escala, sobre todo en cuanto a ingreso o egreso se refiere, baste con investigar la oferta de cursos de preparación para pruebas a gran escala (en especial para el ingreso a instituciones de educación superior, pero también para pruebas con fines de acreditación de conocimientos, las cuales en México son desarrolladas y operadas por el Ceneval); si bien no hay un registro de las escuelas enfocadas a preparar a los estudiantes para estas pruebas, las ofertas en Internet se dan por cientos, incluyendo formación en línea para ello. Además de este tipo de cursos, dentro de las instituciones educativas se ofrecen talleres y seminarios para la preparación de los estudiantes que presentarán estas pruebas; tal es el caso del Instituto Tecnológico y de Estudios Superiores de Monterrey, institución que dedica un semestre al término de los estudios profesionales de los estudiantes para su preparación para la presentación del EGEL. Éste es un problema más que se genera debido a la consideración de la evaluación como elemento de acreditación/castigo/recompensa y no como factor de retroalimentación/mejora. Cuando la evaluación a gran escala se utiliza de esta manera indica que no hay un énfasis en la calidad de la educación, de esta forma se va relegando así a la evaluación a un fin netamente administrativo, usándola como medio de control y fiscalización con el fin de distribución de recompensas y sanciones (Moreno, 2010).

En este mismo sentido, Volante (2006) presenta un resumen de las mayores críticas de la evaluación a gran escala en los últimos años; entre ellas se encuentran las que a continuación se mencionan. Kane (2002) y Popham (2001) –ambos citados en Volante (2006)– afirman que las pruebas de este tipo pueden tener limitaciones psicométricas importantes y aun así ser la base de políticas educativas; y es que hay que reafirmar el hecho de que no hay prueba perfecta, siempre habrá grados de error de medida, errores en el levantamiento de datos o incluso en calificación. Según la Canadian Teachers Federation (2009), la evaluación a gran escala actúa como una forma de control que va en contra de la autonomía de los educadores, ya que al estar alejada de la enseñanza la obliga a centrarse únicamente en los contenidos de las pruebas (como en el

fenómeno de enseñanza para la prueba). Por su parte, la American Education Research Association (2000) y Darling-Hammond (2004) critican el hecho de que la evaluación a gran escala puede orillar a que se mantenga a los docentes como responsables de los resultados, ello a partir de considerar el resultado de los estudiantes como único indicador de la calidad de programas educativos, y como consecuencia, a contribuir a que tengan sentimientos de baja eficacia. McNeil (2000) y Valencia y Villareal (2003) aseguran que la evaluación a gran escala aumenta brechas entre clases sociales ignorando elementos importantes que tienen gran influencia dentro de resultados educativos, como el nivel socioeconómico, lengua de origen, así como su estado físico o emocional, tratando las condiciones de todos los estudiantes de la misma manera. Por otra parte, un pensamiento que comparten Anderson y Postl (2000), Glaser, Linn y Bhornstedt (1997), Hess y Bringham (2000), así como Wideen, O'Shea, Pye e Ivany (1997) es que este tipo de evaluación limita a la educación penalizando el pensamiento crítico y creativo, así como el trabajo intelectual en general, dado que los estándares de evaluación los excluyen como parte de los elementos de análisis. También puede considerarse que la evaluación a gran escala modifica el currículum provocando que se enseñen estrategias orientadas solo a responder adecuadamente las pruebas, lo cual limita el tiempo a otras habilidades que no son evaluadas, en particular cuando los resultados son de alto impacto o de índole únicamente sumativa (Kohn, 2002; Smith y Fey, 2000 y Volante, 2004). Asimismo, si la evaluación a gran escala solo tiene fines de ranqueo o de certificación, la información que proveen resultará de poca utilidad para la práctica en el aula (Stiggins, 2002 y Falk, 1998) e incluso en algunos casos dañar el autoconcepto en los evaluados (Meaghan y Casas, 2001 y Miller y Tovey, 1996).

Por su parte, Moreno (2010) afirma que la evaluación a gran escala es de gran utilidad para la búsqueda de la calidad en la formación, en particular cuando se enfoca *para el aprendizaje*, resaltando las ventajas de que los alumnos sigan aprendiendo y manteniéndolos confiados; sin embargo, este autor coincide con que se corre el riesgo de utilizarla únicamente con fines de control, como se ha comentado. En ese mismo sentido, se debe evitar que el uso de control de la evaluación orille a que se desarrolle lo que Calero y Choi (2012) definen como *el efecto lecho (o cama) de Procusto*<sup>16</sup>, en el cual cuando dentro de la institución y gestión de los programas

---

<sup>16</sup> El uso de la frase “lecho de Procusto” refiere a cuando se acomodan las cosas o datos a la visión u opinión de las personas responsables de su manejo. Su uso se generó de la mitología griega, específicamente de la historia de Procusto, hijo de Poseidón, quien daba estancia a huéspedes para posteriormente tortúralos y modificar sus cuerpos para que éstos se acomodaran al tamaño de sus camas.

educativos se hacen esfuerzos para ajustar poco a poco ciertas condiciones con el fin de alcanzar de forma artificial o forzada los estándares establecidos por las evaluaciones, de esta forma se hace a un lado la idea original de retroalimentación, que es el eje de la evaluación.

Todas estas perspectivas ponen de manifiesto la existencia una fuerte tendencia de crítica a la evaluación a gran escala, y es que cuando se concibe al aprendizaje solamente como un constructo que se puede decantar en mínimas expresiones, la única opción viable para su observación es el uso de pruebas objetivas (Martínez, Cepeda, Fuentes y Burgos, 1995). La evaluación debe considerar entonces la vinculación entre el discurso científico, el objeto y la matriz social dentro de la cual se realiza. Debe buscar conocer las características del aprendizaje y dar una explicación del por qué se da de esa forma. La cuestión se agrava cuando incorporamos estas prácticas a la evaluación del egreso de la formación profesional, y es que “al volver de alto impacto pruebas concebidas con propósitos fundamentalmente formativos se corre un riesgo muy grande de que se produzcan consecuencias indeseables, que ya han comenzado a presentarse, como el que los maestros enseñen para la pruebas o que algunas escuelas usen los resultados para campañas mercadotécnicas sin sustento sólido” (Martínez y Santos, 2009, pág. 273). Debe cuidarse que el contenido de las pruebas a gran escala sirva de eje para la actividad educativa con fines de enfocarse en el aprendizaje.

En este mismo sentido, la apreciación de la AERA (2000) y de Darling-Hammond (2004), ambos citados en Volante (2006), es que uno de los mayores problemas de la evaluación a gran escala pareciera ser la imposición de repercusiones a partir de los resultados obtenidos, ya que puede tener un efecto adverso para la educación, problemática que se denota en las evaluaciones del egreso en México, las cuales condicionan la terminación de estudios profesionales, en algunos casos son utilizadas como forma de titulación, o bien, como condicionante al control de recursos para las IES y los programas educativos. Paradójicamente, como se comentó en el capítulo previo, la búsqueda de repercusiones fueron las que potencializaron el desarrollo de este tipo de herramientas. En el caso del EGEL del Ceneval, se desarrolló inicialmente bajo una propuesta de ANUIES de retroalimentar la eficacia de los programas educativos en cada IES con el objetivo último de mejorar las acciones educativas a nivel superior (Ceneval, 2013a), ello bajo la intención de generar evaluaciones de logros de los estudiantes con carácter de externo a la instituciones educativas, lo cual daría mayor validez y confiabilidad a la información generada. Ahora bien, a partir de esta iniciativa nacida en 1994, las IES mexicanas comenzaron a hacer uso del EGEL como

forma de titulación, considerándolo como sustitución de la opción de titulación comúnmente conocida dentro de las instituciones como *Examen General de Conocimientos*, la cual ha logrado ser una salida de obtención de un título profesional para los egresados que no desean realizar una tesis o un proyecto de investigación, el cual era el modelo tradicional de titulación en México. El EGEL entonces representa una forma efectiva de que los estudiantes puedan obtener su título profesional, reduciendo esfuerzos por parte de las instituciones para la elaboración de una prueba que pretenda cubrir la mayor parte de los conocimientos y habilidades requeridos al finalizar la formación profesional y favoreciendo al mismo tiempo el conocimiento sobre si los egresados alcanzan los estándares fijados a nivel nacional para la disciplina en cuestión. Si bien dentro del Ceneval no existen publicaciones relacionadas con el uso de sus instrumentos, basta con revisar los reglamentos de titulación de diversas IES públicas y privadas que hacen uso del EGEL como forma de titulación, tal es el caso de: Universidad Autónoma del Estado de Morelos, Universidad de Guadalajara, Universidad del Valle de México, Universidad Tecnológica de México, Universidad Autónoma de Chiapas, Universidad Autónoma Indígena de México, Universidad Autónoma del Estado de México, Universidad de Occidente, Universidad Politécnica de Zacatecas, Universidad Autónoma de Zacatecas, Universidad Autónoma del Carmen, Universidad Autónoma de Tamaulipas, Universidad Juárez del Estado de Durango, entre otras. Adicionalmente, hay instituciones que lo utilizan como herramienta de diagnóstico sobre los logros alcanzados por todos sus egresados, estableciéndolo como un requisito para el egreso, tal es el caso de: Universidad Autónoma de Campeche, Universidad Autónoma de Baja California, Universidad Autónoma de Aguascalientes, Universidad Autónoma del Estado de Hidalgo, Instituto Tecnológico y de Estudios Superiores de Monterrey, entre otros.

Adicional al problema ya mencionado sobre usar de forma sumativa los resultados de una evaluación pensada con fines formativos (lo cual es una tendencia evidentemente establecida en las IES mexicanas), ninguna evaluación de este tipo debe sustituir el juicio profesional docente respecto a la graduación de un alumno. Como consecuencia, el reto de la evaluación a gran escala consiste en que los *stakeholders* puedan retomar la información que se genere para fomentar mejoras educativas. Además, si bien el uso de la evaluación a gran escala como forma de afianzamiento del concepto de rendición de cuentas no puede frenarse (Moreno, 2010), su uso formativo enfocado en el aprendizaje no debe ser excluido.

Los problemas descritos, así como la gama de modelos de evaluación que pueden dar sustento a las metodologías que se desarrollan con el fin de retroalimentar la práctica educativa mostrados con anterioridad, permiten valorar la necesidad de repensar la forma en que evaluación a gran escala que se ha utilizado; y es que los problemas actuales de este tipo de evaluación pueden resumirse de la siguiente manera: 1) la falta de identificación de las formas de evaluación a gran escala del egreso respecto al modelo de evaluación que sustenta su desarrollo y uso, y no es que esta falta de explicitación sea dañina, sino que no hay claridad respecto a si los organismos evaluadores tienen certeza del enfoque que sustenta su actividad. 2) La consideración única del modelo tyleriano como forma para evaluar a gran escala el egreso de la formación profesional, lo que conlleva a retroalimentar a los planes de estudios profesionales considerando solo los resultados de los estudiantes en una prueba como único elemento de calidad. 3) Hacer un uso excesivo de la evaluación a gran escala del egreso como herramienta de toma de decisiones de alto impacto o de acreditación, tal y como sucede cuando se considera una evaluación con fines formativos como un elemento de decisión de egreso y/o titulación, medida que modifica el sentido tyleriano de retroalimentación de la formación para convertir a este tipo de evaluación en un premio o ranqueo. 4) Limitar a que la evaluación del egreso se centre solamente en la revisión de logros disciplinares y no a elementos que no sean producto del análisis, creatividad y reflexión de los estudiantes. 5) Que este tipo de evaluación, en potencia de gran relevancia para la retroalimentación de la formación terciaria, se confunda con solo llevar a cabo mediciones del aprendizaje.

Frente a este problema se presenta la necesidad de valorar la posibilidad de realizar la evaluación a gran escala del egreso de la licenciatura bajo un enfoque diferente. Para considerar un modelo de trabajo mayormente enfocado en la retroalimentación, un modelo *comprensivo* a gran escala permitiría no solo entender de mejor manera el sistema, sino a las entidades e incluso a nivel de aula, con los fines de: mejorar la educación, determinar el logro de estudiantes y dar retroalimentación relevante a los *stakeholders* (estudiantes, profesores, tomadores de decisiones y la comunidad en general) (DePascale, 2003). Como se comentó en el Capítulo 1, la evaluación comprensiva propuesta por Stake permitiría acercarse al evaluando, en este caso, al egreso de la licenciatura, considerando aspectos que actualmente no son tomados en cuenta en esta práctica, como: 1) centrarse en las necesidades de los *stakeholders* involucrados e interesados en este tipo de evaluación, lo cual abarca a las IES (que implica la consideración de los diversos programas de

estudio y los diferentes modelos pedagógicos que los sustentan), empleadores, docentes, y, por supuesto, estudiantes y egresados; así como recopilar la información de todos ellos para sustentar de mejor manera la toma de decisiones para la mejora de los programas educativos de formación profesional. 2) No solo se centraría en la observación de logros educativos<sup>17</sup>, sino también en efectos secundarios que no necesariamente están planteados de forma explícita en los objetivos de los programas educativos. 3) Permitiría una reflexión sobre la base filosófica de existir de los programas educativos, considerando a sus egresados como reflejo de la formación otorgada a lo largo de 4 o 5 años, ello haciendo uso de más elementos que solo los resultados cognitivos que los egresados puedan alcanzar. 4) Este enfoque reconocería una metodología orientada hacia las características del evaluando, considerándolo como único, enfrentándose así a la realidad de que el egreso de la licenciatura implica no olvidarse de experiencias humanas, lo que facilita su comprensión de forma contextualizada bajo un momento en particular. 5) Favorecería el uso de los métodos mixtos, valorando la posibilidad de incluir no solo las mediciones cuantitativas, sino también otros elementos sensibles a las particularidades de los evaluados.

Si bien la posibilidad de realizar una evaluación a gran escala del egreso de la licenciatura bajo un modelo comprensivo parece importante para ampliar y mejorar el valioso trabajo que en México se ha logrado en este tema, cabe plantear la pregunta sobre cómo desarrollar una evaluación a gran escala del egreso de la licenciatura que considere los elementos que caracterizan a este modelo de evaluación y que potencializan que los resultados sean una herramienta que fomente su uso pedagógico con la idea de retroalimentar a los programas educativos y así fortalecer la formación profesional en México.

Con fines de dar una respuesta a esta incógnita, el presente trabajo se apoya en el estudio de casos como instrumento de análisis (Stake, 2005; Peña, 2009). La finalidad del estudio de caso es que, a diferencia de otros métodos de investigación, no pretende la generalización, sino abordar a profundidad un fenómeno de interés, con fines de analizar sus particularidades. Es así entonces que el objetivo de esta investigación no irá centrado en la descripción de *la realidad* evaluativa, sino en la construcción de *una realidad* evaluativa o de un *enfoque* de evaluación que sea acorde con un modelo comprensivo, característica distintiva de la metodología de estudio de casos.

---

<sup>17</sup> De hecho, como se ha comentado, este modelo no niega el valor de este indicador, pero no lo considera como elemento único de juicio.

Para apoyar la investigación en cuestión, se analizará el EGEL del Ceneval, ello debido a que representa uno de los mayores esfuerzos a nivel mundial de evaluar el egreso de la licenciatura desde una perspectiva a gran escala, y es, además, el único esfuerzo en México de hacer este tipo de evaluación de una manera sistematizada y con intenciones de ser aplicada en todos los programas educativos del país o a la gran mayoría de ellos. En este sentido, el presente trabajo desarrolla una evaluación del EGEL, una metaevaluación, con fines de desarrollar las bases de una propuesta de evaluación a gran escala bajo una perspectiva comprensiva. Es así como la evaluación del EGEL puede considerarse como un estudio de caso, ya que, de acuerdo con Stake (2005), todos los estudios de evaluación son estudios de caso.

La metodología basada en el estudio de casos tiene como principal objetivo el conocimiento a profundidad de un elemento o fenómeno en particular (Bonafé, 1998). Este tipo de investigación dista de los métodos que hacen uso de técnicas cuantitativas en el sentido que implica que el investigador se acerque a su objetivo de una manera poco ajena al objeto, es decir, requiere de una observación e interpretación por parte del investigador.

Regularmente este tipo de estudios comprenden una fase inicial de preconcepción epistemológica del problema, que posteriormente se contrasta con el trabajo de campo, (la cual puede considerar observación, entrevista o recolección documental). Además, esta metodología se centra en elementos particulares o *micro*, sin que ello quiera decir que se pierda de vista la relación del fenómeno con la estructura de la sociedad y del sistema educativo. Asimismo, reconoce la complejidad del objeto de estudio y se centra en el análisis de aspectos prácticos y situacionales, pero sobre todo, permite la comprensión de significados de contexto de los elementos educativos que analiza (Bonafé, 1998).

Cabe precisar que, de acuerdo con Martínez (2006), una de las principales diferencias de esta forma de trabajo con el método cuantitativo es que en este último se requiere de un planteamiento teórico construido previo que sustente las mediciones y análisis que se efectúen durante la investigación; en cambio, en la metodología cualitativa, de la cual el estudio de casos forma parte, las teorías se construyen a partir de los elementos extraídos de un cuerpo teórico que sirve como eje para el investigador, del cual no es necesario obtener una muestra representativa. Frente a ello, Glaser y Strauss (1967, en Martínez, 2006) hacen hincapié en establecer una distinción clara entre la investigación centrada en probar teorías, así como de

aquella enfocada en la generación de teorías. La primera de ellas tiende a tener elementos más cercanos a la metodología cuantitativa, ya que permite robustecer su aplicabilidad. Por su parte, Sarabia (1999, en Martínez, 2006) considera que la investigación es un *espiral inductivo-hipotético-deductivo*, el cual conlleva dos procesos esenciales: la fase *heurística* de observación, descripción y generalización inductiva, así como la *confirmativa*, en la cual se pueden comprobar los elementos de hipótesis desarrollado en la fase heurística. Por ende, la investigación debe considerar acciones tales como la observación, la exploración de la realidad con el fin de desarrollo de hipótesis sobre el fenómeno de interés, así como la contrastación de dicha hipótesis para verificar una posible explicación del fenómeno. Bajo ese contexto, normalmente las metodologías cualitativas se centran en la fase heurística de los fenómenos.

Es necesario considerar que la metodología de estudio de casos es más acorde con una perspectiva centrada en la generación de teorías, su finalidad no es la generalización de conclusiones, aunque sí puede valorar la posibilidad de transferibilidad entre teorías (Maxwell, 1998). Por tanto, si esta metodología pretende el desarrollo de una teoría, es necesario que cuente con elementos que permitan asegurar su validez interna, externa, así como su confiabilidad.

De acuerdo con Yin (1989, en Martínez, 2006), en la metodología de estudio de caso el investigador debe tener claridad respecto a las proposiciones teóricas que establecen las variables o constructos que es necesario considerar dentro del fenómeno por investigar. Dentro de las proposiciones teóricas se deben incluir las unidades de análisis (si es un caso único o varios, o bien, un caso con subcomponentes que sea necesario analizar) y las fuentes de las cuales se obtendrá información para su consideración. Adicionalmente, parte del trabajo del investigador es la vinculación lógica de los datos con las proposiciones teóricas, estableciendo de forma explícita los criterios para la interpretación de los datos. Finalmente, los resultados de la investigación se presentan mediante las conclusiones, las cuales fortalecen las teorías que enmarcan la investigación; esto quiere decir que para el desarrollo del estudio de caso es necesario realizar un protocolo de las tareas por desarrollar, especificando una semblanza del caso, las preguntas del estudio por realizar y los procedimientos por llevar a cabo.

Bajo esta lógica metodológica, la forma de abordar la construcción de los elementos de la evaluación a gran escala del egreso de la licenciatura bajo una perspectiva comprensiva implica la

descripción a profundidad del EGEL acompañada con una crítica bajo el enfoque comprensivo de evaluación, lo cual permitirá avanzar en los establecimientos metodológicos y operativos para alcanzar el objetivo propuesto. De esta forma, siguiendo la forma de trabajo de Schutz (1974), la realidad por analizar (en este caso, el EGEL) no tiene un sentido en sí mismo, dado que el fenómeno ha sido diseñado por los actores involucrados; por lo que las pautas típicas de acción crean un mundo de expectativas apoyadas socialmente, estas expectativas servirán para la comprensión del fenómeno en particular y así poder realizar un planteamiento metodológico distinto al observado. Para ello, la descripción del EGEL se realizará considerando su metodología de desarrollo sin centrarse en particularidades de alguna de las 40 disciplinas a las cuales está dirigido, ello debido a que cada EGEL se elabora e implementa con las mismas reglas metodológicas. Considerando esta situación, los elementos por describir y analizar para efectos de la construcción de un análisis de caso, así como las preguntas que orientan la descripción de cada aspecto son los siguientes:

- Inicios. *¿Cuáles fueron los acontecimientos y/o necesidades que llevaron a la ANUIES a la creación de los EGEL?*
- Evolución. *¿Cuál es el estatus actual de los EGEL y cuáles son los aspectos sociales y académicos que han favorecido su desarrollo?*
- Concepción teórica del EGEL. *¿Cuál es el modelo que sustenta la actividad evaluativa de los EGEL?*
- Metodología para su elaboración
  - a. Definición teórica del constructo de medida. *¿Cuál es la metodología que permite la definición teórica del constructo de medición?*
  - b. Elaboración de especificaciones. *¿Cómo es el proceso de desarrollo de especificaciones por evaluar?*
  - c. Elaboración de reactivos. *¿Cuál es el procedimiento que se sigue para el desarrollo de reactivos?*
- Proceso de aplicación. *¿Cómo se desarrolla el proceso de aplicación?*
- Proceso de definición de puntos de corte y calificación. *¿Qué metodología se utiliza para la definición de puntos de corte y emisión de calificación?*
- Emisión de resultados. *¿Cuáles son los elementos de decisión que soportan la forma de presentar resultados?*

Los elementos enlistados se consideraron a partir de aspectos definitorios al EGEL, con base en los siguientes documentos publicados por el Ceneval entre 2012 y 2013: *Manual técnico del EGEL*; *Metodología Ceneval*; *Encuesta nacional para la validación social de los perfiles profesionales de los EGEL*; así como el *Reglamento de los consejos técnicos de los exámenes*. Dentro de estos documentos el Ceneval hace una descripción de las características del EGEL y de los órganos colegiados de los cuales se apoya para su elaboración. Considerando estos insumos, así como mi experiencia en la coordinación de tres EGEL (Comercio/Negocios Internacionales, Turismo y Gastronomía) desde el año 2009 al 2013, se realizará una descripción detallada de los seis elementos previamente mencionados. La conjugación de estos elementos me permitirá la construcción de un estudio de casos detallado. Una descripción a profundidad de todos los aspectos de esta prueba facilitará la elaboración de un juicio de valor sobre cada elemento asociado al desarrollo e implementación de un EGEL, así como orientar la definición de una propuesta metodológica de este tipo de evaluación bajo una metodología comprensiva. Este análisis se desarrolla en el siguiente capítulo, el cual sirve de insumo principal para la propuesta planteada en el Capítulo 5.

## 4. Análisis de caso: EGEL

En capítulos previos se ha abordado la necesidad de reflexionar sobre el modelo de evaluación que ha definido a la evaluación a gran escala. Es a partir de esa necesidad que a continuación se presenta un análisis o estudio de caso sobre el Examen General para el Egreso de la Licenciatura (EGEL) del Ceneval como elemento eje para el desarrollo de una propuesta de evaluación a gran escala bajo un modelo comprensivo.

La selección del EGEL como elemento de análisis y eje de una propuesta comprensiva responde al hecho de que representa uno de los mayores esfuerzos de evaluación a gran escala en México, además que es el proyecto más grande de evaluación del egreso de la licenciatura en nuestro país y está a la par de otros ejemplos de este tipo de evaluación en diversos países (como se comentó en el Capítulo 2). El EGEL sin duda representa un gran esfuerzo de diversas IES que no debe dejarse a un lado debido a la forma en que el proyecto ha enfrentado la valoración de su egreso para su inserción en el ámbito profesional, así como un esfuerzo interinstitucional de consolidar proyectos en común; este esfuerzo representa entonces un soporte importante para el desarrollo de nuevas propuestas de evaluación del egreso de la licenciatura a gran escala.

Considerando la relevancia del EGEL como aspecto histórico en la evaluación del egreso de la licenciatura a gran escala, el presente capítulo se centra en la descripción a profundidad de esta evaluación, desde sus antecedentes, su evolución, concepción teórica, metodología de elaboración, proceso de aplicación y calificación, así como la emisión de resultados. Como se mencionó, los insumos principales son los documentos liberados por el Ceneval, así como mi experiencia de trabajo como coordinador de tres diversos EGEL (Comercio/Negocios Internacionales, Turismo y Gastronomía) del año 2009 al 2013.

### A manera de introducción: ¿qué es el EGEL?

Los EGEL “son instrumentos de evaluación de cobertura nacional que evalúan el nivel de conocimientos y habilidades académicas de los recién egresados de una carrera universitaria” (Ceneval, 2003, pág. 5). Son instrumentos que están desarrollados para poder identificar si los egresados de nivel profesional tienen los conocimientos y habilidades mínimos requeridos para comenzar de forma eficaz su ejercicio profesional. Al año 2014 existían 40 EGEL, cada uno de ellos

centrado en la evaluación del egreso de una disciplina o profesión en particular, a saber: Administración, Arquitectura, Biología, Ciencia Política y Administración Pública, Ciencias agrícolas, Ciencias computacionales, Ciencias de la comunicación, Comercio/Negocios Internacionales, Contaduría, Derecho, Diseño gráfico, Economía, Enfermería, Gastronomía, Informática, Ingeniería Civil, Ingeniería Computacional, Ingeniería de Software, Ingeniería Eléctrica, Ingeniería Electrónica, Ingeniería en Alimentos, Ingeniería Industrial, Ingeniería Mecánica, Ingeniería Mecánica Eléctrica, Ingeniería Mecatrónica, Ingeniería Química, Medicina General, Medicina Veterinaria y Zootecnia, Mercadotecnia, Nutrición, Odontología, Pedagogía-Ciencias de la Educación, Psicología, Química, Química Clínica, Químico Farmacéutico Biólogo, Relaciones Internacionales, Trabajo Social y Turismo.

Cada EGEL es una prueba de opción múltiple que se aplica en lápiz y papel o en línea cuyo número de reactivos varía conforme a la estructura o dominios de cada disciplina que se pretende medir; el único EGEL que considera la ejecución de los evaluados de una forma distinta es Arquitectura, el cual requiere el desarrollo de un proyecto arquitectónico que hace uso del juicio humano para su valoración. El desarrollo de los 40 EGEL, al menos en cuanto a lo que refiere a la estructura basada en reactivos de opción múltiple, mantiene una forma única de elaboración. Esta estructura le permite contar con las siguientes características (Ceneval, 2013b):

- Especializado por carrera profesional. Como se comentó, cada EGEL pretende evaluar conocimientos y habilidades particulares de cada disciplina o profesión, no se centra en habilidades genéricas, como la lectura, escritura o tipo de razonamiento.
- De alcance nacional. Implica que su desarrollo le permite tener validez en todo el país, ya que no se fundamenta en ningún currículo en particular.
- Estandarizado. Debido a la necesidad de mantener un instrumento cuya esencia sea fija, el EGEL tiene reglas determinadas para su diseño, elaboración, aplicación y calificación.
- Criterial. A diferencia de otro tipo de evaluaciones que basan sus formas de calificación a partir de la norma, el EGEL se considera criterial, dado que los resultados de cada evaluado o sustentante se comparan con un estándar nacional de desempeño previamente establecido; ello permite que los resultados de los sustentantes no estén en función de las condiciones de evaluación o de los resultados de los sustentantes contra los que se compite, como suele suceder en exámenes de ingreso.

- De máximo esfuerzo. Se infiere que el nivel de rendimiento mostrado por una persona se sustenta en su mejor esfuerzo para responder a la evaluación.
- De alto impacto. A partir de los resultados obtenidos las IES obtienen un indicador que les permitirá definir acciones por seguir en la trayectoria de la persona, o bien, tomar decisiones sobre el programa educativo.
- De opción múltiple. Cada reactivo incluido en la prueba contiene cuatro opciones de respuesta de las cuales solo una es la correcta.
- Contenidos centrados en problemas. El desarrollo de reactivos mantiene una base problemática, no centrada en lo conceptual, ello con la intención de determinar si las personas que lo sustentan son capaces de utilizar lo que han aprendido en la licenciatura.
- Sensible a la instrucción. Debido a que evalúa resultados de aprendizaje de una formación profesional, se infiere que éstos son el resultado de una experiencia educativa dentro de una IES.
- Contenidos validados socialmente. La perspectiva de los profesionales en ejercicio que participan en el desarrollo de la prueba favorece a su validez social.

El desarrollo de cada EGEL, así como el trabajo que se realiza de forma cotidiana para su mantenimiento se encamina para cumplir con los elementos mencionados. Esta breve descripción del EGEL nos permite valorar en primera instancia su objetivo predominantemente técnico, sin embargo, valdría conocer más a fondo esta evaluación a fin de identificar con mayor precisión su enfoque de evaluación. Frente a ello y con la intención de mantener consistencia respecto a la consideración de un modelo comprensivo, resulta pertinente entonces mostrar una descripción profunda del EGEL desde sus inicios.

### **Inicios. ¿Cuáles fueron los acontecimientos y/o necesidades que llevaron a la ANUIES a la creación de los EGEL?**

Ceneval es una asociación civil mexicana sin fines de lucro creada en el año de 1994. En 1993 la Asamblea General de la ANUIES turnó a la Secretaría de Educación Pública la propuesta de creación de un “organismo no gubernamental autofinanciable a mediano plazo que se ocupara de

llevar a cabo el conjunto de acciones necesarias para realizar los siguientes exámenes de cobertura nacional:

- Para evaluar a los estudiantes que egresan de la educación secundaria y solicitan ingreso al bachillerato y otras modalidades de educación media superior...
- Para evaluar a los estudiantes que solicitan ingreso a la licenciatura...
- Examen General de Calidad Profesional..." (Gago, 1995; pág. 1)

A partir del 28 de abril de 1994 el Ceneval obtuvo personalidad jurídica para desempeñar las funciones encomendadas por la ANUIES con el apoyo del presupuesto de dicha organización. El primer antecedente del EGEL fue el Examen General de Calidad Profesional (EGCP), que fue como se les denominó en sus inicios. Los primeros EGCP fueron Contaduría y Medicina Veterinaria y Zootecnia, mientras que en 1995 comenzaron los trabajos para Enfermería, Medicina, Odontología y Químico Farmacéutico-Biólogo.

De acuerdo con Antonio Gago (2000), primer director del Ceneval, la razón principal para la creación del Ceneval fue la inconformidad sobre la calidad de la educación superior manifestada particularmente por los rectores de las IES mexicanas. Frente a una demanda de la educación superior para mejorar su calidad, la ANUIES formuló la propuesta de un organismo externo a las instituciones que se encargara de evaluar los programas educativos desde una perspectiva ajena a las mismas. Como se abordó en el capítulo previo, la instauración de organismos dedicados a la evaluación (entre ellos Ceneval) facilitaría la distribución del financiamiento a la educación superior (Villaseñor, 2003). La creación del Ceneval como instancia de evaluación externa de los procesos educativos se dio en un momento en el cual en diversos países de Latinoamérica crearon organismos o agencias para la evaluación y acreditación de programas de educación superior (entre ellos Argentina, Chile, Brasil, Paraguay, Bolivia, Ecuador, Colombia, Costa Rica, El Salvador, México, Cuba, Trinidad y Tobago, así como en Jamaica) (De la Garza, 2008).

La búsqueda de la mejora de la calidad educativa a través de su evaluación responde a una dinámica internacional de libre mercado que tuvo su auge en los noventa, donde el factor competencia es el eje para ofertar opciones al consumidor (Viñao, 2012). En ese sentido, la búsqueda de una educación superior de calidad favorecería a la competencia económica del país; y, como se ha comentado ya, bajo este contexto, la evaluación a gran escala permitiría la

implementación de esta política. En ese sentido, desde sus inicios, la línea de trabajo del Ceneval se centró en la evaluación del logro de los estudiantes, ya que consideraba que “la calidad de un programa educativo se mide, entre otras cosas, a partir del aprendizaje que logran los estudiantes” (Gago, pág. 5).

La formulación de los EGCP se remonta a marzo de 1992, donde en la reunión de rectores de las universidades mexicanas, el Secretario de Educación Pública (SEP) en aquel entonces, Ernesto Zedillo Ponce de León, expuso la necesidad de continuar con el desarrollo del Sistema Nacional de Evaluación de Educación Superior mediante la generación de instrumentos centrados en la evaluación de resultados. La importancia de este tipo de evaluaciones tiene una trascendencia de Estado, la cual se dio bajo el contexto de un año importante para la política educativa, en particular debido al Acuerdo Nacional para la Modernización de la Educación Básica y Normal, que fuera publicada en el Diario Oficial de la Federación en mayo de 1992. El Acuerdo conllevó cambios en la perspectiva de la educación básica que aún persisten, pero, adicionalmente, la perspectiva de modernización y procuración de la calidad educativa también se vio reflejada en la educación superior. Al año 2014, el Sistema Nacional de Evaluación de Educación Superior consideraba una estructura que soporta decisiones de alto impacto en la vida de las IES, al estar conformado por el Consejo de Acreditación de la Educación Superior, A.C. (COPAES), el otorgamiento del Registro de Validez Oficial de Estudios (RVOE) de la SEP, el Programa Integral de Fortalecimiento Institucional (PIFI), el Programa para el Mejoramiento del Profesorado (PROMEP), el Sistema Nacional de Investigadores (SNI) del Consejo Nacional de Ciencia y Tecnología (Conacyt), la Dirección General de Profesiones (DGP) de la SEP, así como el Ceneval.

Como se mencionó en el Capítulo 2, en México, a partir de los años ochenta, a pesar de los problemas económicos que se suscitaban en el país, comenzó un proceso social de masificación de la educación (Acosta, 1995), así como la necesidad de atender las demandas de competitividad, las cuales presionaban a las IES en varios sentidos, a saber: 1) en la necesidad de atención de una creciente demanda estudiantil (centrada en la intención de una mejora social frente a un contexto mexicano cuya economía se presentaba en desventaja bajo la dinámica globalizada en la cual se veía inmerso el país); 2) en asegurarse del cumplimiento de su función social de formación profesional bajo una política de financiamiento de la educación superior coordinada mediante el Programa Nacional de Educación Superior (desarrollado por la CONPES), y en un segundo momento por el Programa Integral para el Desarrollo de la Educación Superior, aprobado por

CONPES. Es a partir de estas regulaciones que se insta el otorgamiento de financiamiento en función de la evaluación de resultados (Rodríguez, 2002); 3) centrar sus procesos educativos en alcanzar una calidad, buscando un el reconocimiento público de ésta y servicio a la sociedad (Casillas y De Garay, 1993). De esta forma, es a partir de los ochenta y de la política educativa de Miguel de la Madrid que comienza el interés en evaluación como elemento al servicio del Estado, ya que permitiría identificar problemas y jerarquizar las necesidades específicas de los ámbitos educativos (Martínez, Cepeda, Fuentes, y Burgos, 1995); asimismo, como consecuencia de la crisis económica de los ochenta, se permitió la modificación del esquema de financiamiento de las IES públicas, dado que, bajo una perspectiva de rendición de cuentas, el financiamiento público a la educación superior no podía mantenerse de la misma manera (Márquez, 2004). Con este antecedente, la evaluación tiene entonces un papel crucial en el desarrollo de la política educativa mexicana, que si bien en el discurso oficial se concibe con la función de retroalimentación del sistema, desde la creación del Ceneval las instituciones han adoptado el modelo de acreditación dejando atrás la evaluación con fines de diagnóstico y mejora (Rodríguez, 2002).

Este contexto político dentro del cual el Estado se interesaba en la calidad de la educación frente a una nueva dinámica educativa enfocada a la movilidad social y a la obtención de recursos para hacer frente a nuevas demandas de calidad educativa, en particular en educación superior, contribuyó a que en los años noventa la política de la SEP considerara la integración de evaluaciones dentro de la agenda educativa; y, si bien no se contemplaba aún al Ceneval como órgano operativo encargado de la evaluación de los resultados de los estudiantes, la conformación del Sistema Nacional de Evaluación de Educación Superior era prueba fehaciente de ello. Como se comentó previamente, la instancia encargada del Sistema Nacional de Evaluación de Educación Superior fue la Comisión Nacional de Evaluación (CONAEVA), fundada en 1989 e integrada por representantes del Gobierno Federal y cuatro miembros del Consejo Nacional de la ANUIES (Rubio, 2007). La CONAEVA consideró la creación de unidades de evaluación dentro de las IES, de organismos de evaluación diagnóstica, de evaluación de los aprendizajes de los estudiantes, así como de agencias acreditadoras de los programas educativos. Bajo este contexto, correspondería más adelante al Ceneval la ejecución de la evaluación de los aprendizajes bajo los lineamientos de la CONAEVA. De esta forma, como lo comenta Acosta (1995), la CONAEVA facilitó la implementación de una política de educación superior basada en la evaluación.

En el proceso de consolidación del Sistema Nacional de Evaluación de Educación Superior, posterior a la reunión de la SEP con los rectores en 1992, en la Revista de la Educación Superior de la ANUIES (1993) se publicó la propuesta de creación de la EGCP, misma que había sido presentada en la XXV Sesión Ordinaria de la Asamblea General de la ANUIES. La propuesta inicial consideraba la creación de pruebas centradas en los egresados de instituciones de educación superior del país con el objetivo de medir los conocimientos y habilidades de su formación académica. Este objetivo de medida sigue manteniéndose vigente en la actualidad, aunque también hay que considerar su principal meta:

“Los principales objetivos de estos exámenes son: informar a la sociedad mexicana acerca de la calidad en la formación académica de los nuevos profesionistas, contribuir a la evaluación de la calidad de la educación superior, aportar información que permita definir acciones eficaces para mejorarla, ampliar el conocimiento del que disponen las instancias gubernamentales para establecer políticas de apoyo y asignación de recursos y fundamentalmente informar al propio sustentante acerca del nivel de formación alcanzado. Los exámenes generales desempeñarán funciones esencialmente constructivas y educativas, en lugar de la función meramente calificadora, excluyente y penalizadora que frecuentemente se ha atribuido a la evaluación del conocimiento” (ANUIES, 1993).

La meta de los EGCP era muy clara respecto al sentido de mejora de la educación superior en función de los resultados observados, centrándose sobre todo en funciones constructivas, coincidiendo con ello con la política educativa del Estado a principios de los años noventa, resultantes de los procesos de masificación educativa y calidad (entendida como la búsqueda de la mejora de la competencia) que se vivían en aquel entonces y que se mantiene en la actualidad. Cabe resaltar también la precisión realizada sobre su esencia ajena a la penalización, así como a la exclusión y a la función netamente calificadora, lo cual se destaca de las metas de los EGCP. Si bien en la actualidad existe una cultura de la evaluación que se va afianzando poco a poco dentro de las instituciones dedicadas a la educación (de cualquier nivel educativo), misma que también se refleja en las políticas de Estado (como la Reforma Educativa del 2013, la cual, como se comentó, considera a la evaluación educativa como elemento de gran importancia dentro de la educación pública), a principios de los noventa esto no era así.

La creación del Ceneval se debe a la dinámica educativa de finales de los ochenta y principios de los noventa, pero fue la propuesta de uso de los EGCP como insumo de análisis de la calidad de la educación lo que permitió su consolidación. De esta manera, los EGCP o EGEL representaron una respuesta operativa a una política interesada en la calidad educativa basada en la rendición de

cuentas, considerando a los resultados de los egresados de programas de educación superior como un elemento de la calidad educativa recibida en las IES. De ahí la relevancia que las siglas de los instrumentos incluyeran al término “calidad profesional” (que cuando terminó la administración de Gago dentro del Ceneval se modificó a “egreso de licenciatura”), dado que su nombre reflejaba la política nacional de la importancia de la evaluación en la educación como medio de establecimiento de calidad.

La conformación del Sistema Nacional de Evaluación de Educación Superior y la propuesta de desarrollo de los EGCP dentro de ello, se mantiene en concordancia con la forma de generación del modelo de evaluación centrado en objetivos de Tyler; en el sentido que ambos surgen como una necesidad social y de Estado de conocer los avances en materia educativa en función de procesos de masificación de la educación con fines de movilidad social, así como también de la generación de un indicador que fuera insumo para la rendición de cuentas y dar elementos para la decisión relacionada con apoyos y asignación de recursos.

Esta rápida revisión de los documentos oficiales que describen el contexto político que permitió la generación de los EGEL favorece una comprensión de la intención de su desarrollo, en particular el por qué se enfocó en los resultados de los estudiantes como reflejo de la calidad educativa; y es que esta evaluación se mantiene como un elemento del Sistema Nacional de Evaluación de Educación Superior, el cual requiere considerarse como un abanico de elementos de análisis que tienen por objeto favorecer el avance de la educación superior. Adicionalmente, vale resaltar que el EGEL se concibió inicialmente como una prueba; es decir, no se pensó propiamente como una evaluación. Dicho discurso, el cual se centra propiamente en la medición más que a la retroalimentación o en la indagación del porqué de los resultados educativos, resultaría benéfico para un acercamiento a la búsqueda sobre los avances educativos, y es que una sociedad que en los años noventa no estaba cercana a procesos de evaluación y que resaltaba la intención de ser competitivo en términos económicos, pensar en la evaluación como un sinónimo de la medición del aprendizaje resultaba conveniente. Por supuesto, el discurso oficial del Ceneval respecto al uso de las pruebas como medio de retroalimentación académico no convenció a todos. Por ejemplo, a fines de los años noventa la Universidad Nacional Autónoma de México (UNAM) formaba parte de la Asamblea de Asociados del Ceneval, sin embargo, como producto de la huelga universitaria en

el año de 1999, dejó de participar en dicho cuerpo<sup>18</sup> y las evaluaciones del Ceneval no se aplicaron en dicha institución. En el mismo sentido, Villaseñor (2003) afirma que Ceneval logró la implementación de los EGEL como medio de sustitución de los sistemas de titulación de las IES, atendiendo así contra la autonomía de las instituciones; de esta forma, la evaluación ya no se centraba en la retroalimentación, sino en la búsqueda de acreditación con fines de obtención de recursos concursables.

### **Evolución. ¿Cuál es el estatus actual de los EGEL y cuáles son los aspectos sociales y académicos que han favorecido su desarrollo?**

Los EGCP comenzaron en el año de 1994 con Contaduría y Medicina Veterinaria y Zootecnia, aunque actualmente existen 40 EGEL para diversas licenciaturas. El incremento del número de pruebas por disciplina refleja sin duda un aumento de su demanda por parte de las instituciones educativas. Como lo comenta Gago (2002), esta evaluación comenzó como voluntaria, por lo que el incremento de su demanda reflejada en el número de EGEL disponible para las IES muestra una importante estabilidad, misma que se ha venido estableciendo por la demanda de las instituciones y a pesar de haber surgido en un contexto en el cual había poca información sobre la evaluación en educación. Incluso, de acuerdo con datos del Ceneval (2012), el número de evaluados o sustentantes que participaron en el EGEL de 1994 al 2011 era de 853,000 egresados de licenciatura pertenecientes a más de 400 plantel o campus, lo cual es otro indicador de la solicitud que ha tenido esta evaluación.

Hay que considerar que la enorme demanda del EGEL dentro de las instituciones educativas, como se ha comentado, se da dentro de un contexto de masificación de la educación superior. De acuerdo con Rama (2009), en toda América Latina se ha presentado el fenómeno de masificación de la educación superior desde los años ochenta del siglo anterior. Este proceso no se debe únicamente al conjunto de egresados de nivel bachillerato que desea ocupar un lugar en educación superior, sino que intervienen otros factores tales como la búsqueda de movilidad social mediante la profesionalización, así como la gratuidad de las instituciones educativas e incluso la modificación de la estructura laboral que actualmente demanda con mayor fuerza a profesionales y técnicos superiores; y es que los profesionales universitarios y técnicos tienden a

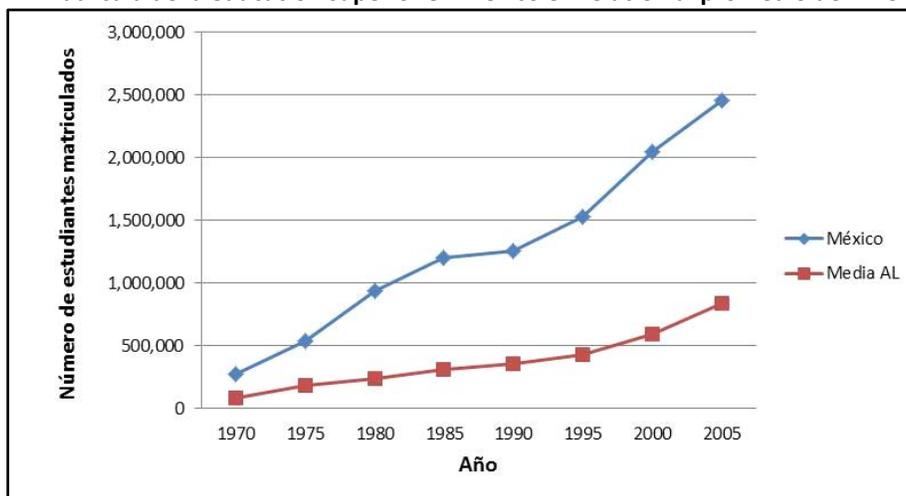
---

<sup>18</sup> Ello debido a que se consideraba que el EGEL violaba la autonomía universitaria en lo relacionado con la gestión que la universidad realiza de sus propios procesos de evaluación.

percibir mayores ingresos salariales y los niveles de desempleo con esos estudios son relativamente menores (Rama, 2009).

En particular en el caso de México, la matrícula en educación superior es muy alta en relación a otros países de América Latina. La Gráfica 4.1 muestra la tendencia de la matrícula en este nivel educativo.

**Gráfica 4.1. Matrícula de la educación superior en México en relación al promedio de América Latina.**



Fuente: Elaboración propia con datos de Rama (2009).

Como se puede observar, la matrícula en educación superior en México es mayor que el promedio de América Latina y ha mostrado una clara tendencia de aumento desde 1970, aunque es después de 1990 que la tendencia de la matrícula muestra una inflexión importante, acelerando su crecimiento hasta años recientes. La educación superior en México desde los años noventa comenzó a ser un tipo de educación de masas y no de una élite social. Adicional a la preocupación de la calidad educativa entre programas educativos, parecía que frente a un cambio en la dinámica de la educación superior orillarían a que las IES valoraran sus mecanismos sobre el control de la demanda, y, por supuesto, de la matrícula que cada año generará una petición de egreso.

Considerando este contexto de alta demanda de la educación superior en México, resulta pertinente recordar que las escuelas, como parte de un proyecto social, implican una gestión y un control (Salinas, 2002); son espacios para el conocimiento del mundo organizados de una manera sistematizada dentro de las instituciones y, como tal, la evaluación es parte de esa gestión, es una actividad regulada e institucionalizada. Valorar a la evaluación como un elemento institucionalizado que participa dentro de la regulación del conocimiento mediante la práctica

organizada permite entonces ampliar la visión sobre la evaluación y su rol en el contexto del aumento de la demanda a nivel superior.

La evaluación de los aprendizajes, en particular la del egreso del nivel superior, es un agente institucional a pesar de que su carácter pueda ser externo, tal como lo es el EGEL. Es menester considerar al EGEL, y a la evaluación en general, como un elemento institucional de observación y regulación. Valorar a la evaluación de forma ajena a los procesos institucionales que le dan cabida dentro de las IES no es adecuado, ya que, como comenta Salinas (2002), si solo se consideran elementos técnicos la evaluación se vuelve sencilla, dado que se deja a un lado su papel social y político. Adicionalmente, se debe considerar que los procesos institucionales de evaluación se dan a la par o como consecuencia de procesos sociales y políticas de un momento histórico en particular.

Bajo esta lógica, la evaluación es “un conjunto de experiencias y vivencias de profesores y alumnos que tienden a tratar de evidenciar o constatar determinados aprendizajes del alumno (unos y otros no) con la finalidad tanto de mejorarlos como de juzgarlos desde un marco institucional que regula, distribuye y valora el conocimiento de dichos juicios como referencia” (Salinas, 2002; pág. 21). Como tal, podremos valorar la implementación del EGEL dentro de la agenda de las IES en México como un elemento consecuente de la política estatal centrada en la calidad y rendición de cuentas, pero además su posicionamiento y evolución puede atribuirse a la presión política de mostrar las evidencias del avance acumulativo del aprendizaje de los egresados como indicador para esta rendición de cuentas; de esta forma, permite dar información para que las instancias financiadoras tengan conocimiento de los avances mostrados por las instancias educativas; ello lleva al concepto de *estado evaluador* (Calero y Choi, 2012). Las IES que optan por la aplicación del EGEL en sus egresados favorecen que los procesos de egreso y titulación se realicen de forma más dinámica, incrementando así los indicadores de ambos aspectos, lo cual puede orillar a la obtención de más recursos y un mejor reconocimiento de sus programas educativos como *licenciaturas de calidad*, ello bajo procesos de evaluación de organismos como COPAES y CIEES, centrados en la búsqueda de acreditación de programas educativos bajo una mirada centrada en la competencia. Las IES también se benefician del uso de los EGEL para incorporarse al Programa Integral de Fortalecimiento Institucional (PIFI) de la SEP (Ceneval, 2012). El PIFI es una estrategia de la SEP para apoyar a las IES a mejorar el nivel de calidad de sus servicios educativos; a través de él las IES reciben recursos para que éstas puedan atender los elementos que se incluyan en su

planeación estratégica (SEP, 2014). El PIFI mantiene la línea política de los programas de los 90, pero también propone integrar los procesos de planeación académica de las instituciones para la obtención de recursos de fondos concursables (Rodríguez, 2002). Actualmente, como lo comenta Villaseñor (2003), gracias al PIFI, los proyectos que se buscan son aquellos que se relacionan con la academia, en particular aquellos considerados como de calidad para que las IES sean acordes con las políticas globalizantes del Estado. Siendo así, esta participación facilita a las IES la obtención de recursos para la presentación del EGEL, lo cual permite que se genere un indicador de resultados de aprendizaje para un programa educativo sin necesidad que los estudiantes desembolsen recursos personales o que las IES destinen sus recursos para ello.

Este acercamiento permite valorar que el desarrollo y el uso del EGEL de forma constante por una gran cantidad de IES en nuestro país están fuertemente ligados a los procesos de rendición de cuentas, sin embargo, también debe reflexionarse su uso como elemento regulatorio de la titulación. La titulación, es decir, la obtención de un grado de estudios, es un elemento diferenciado del egreso, dentro del cual cada IES estipula sus requisitos y formas para la obtención del grado académico, de esta forma los egresados de las instituciones que cuentan con RVOE de la SEP podrán obtener su Cédula Profesional para poder ejercer de forma legal la profesión en nuestro país. Ahora bien, como se comentó en el apartado previo, el EGEL se desarrolló –al menos bajo el discurso oficial– con fines predominantemente de retroalimentación de calidad educativa, aunque su uso se centró prácticamente como una herramienta de titulación para los egresados. De acuerdo con Jorge Hernández, ex Director General Adjunto de los EGEL, en su presentación en la XIII Reunión Ordinaria de la Asociación Nacional de Escuelas y Facultades de Pedagogía y Educación, A.C. (ANEFEP) del 30 de septiembre de 2011:

“...en pocas IES es posible hacer inferencias válidas sobre la calidad de sus programas de licenciatura con base en los resultados del EGEL, ya que NO presenta el EGEL el total de los egresados del programa, ni tampoco, una muestra representativa de ellos; los sustentantes no realizan su máximo esfuerzo al resolverlo; se utiliza el EGEL sólo como una opción de titulación o un simple requisito administrativo, sin consecuencias académicas de ningún tipo” (Hernández, 2011).

El diagnóstico utilizado por el propio Ceneval dio a lugar a la creación del Indicador de Desempeño Académico por Programas de Licenciatura (IDAP), del cual se hablará más adelante. Por ahora baste con identificar el uso del EGEL centrado en una opción de titulación o solo como un requisito administrativo para el egreso o para la titulación misma. Este aspecto denota la relación de la

evaluación como elemento regulador dentro de la dinámica educativa institucionalizada, en el marco en la cual el EGEL ha sido una valiosa herramienta para las IES con una matrícula creciente cada año y con necesidades de titulación y egreso que aumentan ciclo con ciclo; ya que hacer uso de este instrumento desarrollado de forma externa a las IES y que puede ser un referente nacional debido a su metodología de elaboración, favoreció su utilización como elemento regulador del egreso y la titulación frente a una creciente matrícula en un entorno administrativo con pocos recursos financieros disponibles.

Paradójicamente el uso del EGEL que, como se comentó es realmente alto en número de sustentantes y de instituciones usuarias, no se utiliza con los fines de retroalimentación que le dieron vida, sino como una herramienta reguladora del egreso y la titulación; lo cual, no resulta un uso con fines de mejora de la calidad educativa. Adicionalmente preocupan las palabras de los representantes del Ceneval al considerar esta importante herramienta como un elemento que debe ser ligado a consecuencias académicas, faltando así aún más al objetivo primordial de “contribuir a la evaluación de la calidad de la educación superior, aportar información que permita definir acciones eficaces para mejorarla” (ANUIES, 1993) que le dio vida a este proyecto. Este discurso refleja cómo incluso dentro del Ceneval se concibe a esta evaluación como una herramienta de regulación administrativa.

Dentro de la situación actual del EGEL, a finales del año 2010 la Asamblea Directiva del Ceneval aceptó la creación del Programa de Licenciatura de Alto Rendimiento; el cual, mediante el uso del Indicador de Desempeño Académico por Programas de Licenciatura (IDAP), establece de forma anual un padrón de programas educativos cuyos egresados (o una muestra representativa de ellos) presentan buenos resultados en el EGEL. El IDAP, el cual es el indicador para la inclusión de los programas-campus en el padrón, se calcula considerando dos aspectos: a) La proporción de sustentantes a nivel nacional que obtenga como resultado global en el EGEL Testimonio de Desempeño Satisfactorio (TDS), Testimonio de Desempeño Sobresaliente (TDSS) o Sin Testimonio (ST) durante un periodo de referencia (comprendido siempre entre el 1 de julio de un año y el 30 de junio del siguiente), b) La proporción de sustentantes que obtenga TDSS, TDS o ST, en el programa-campus de la IES, en el mismo periodo de referencia. Con base en el resultado alcanzado (cuyos valores estándar oscilan entre -2 a 4), el Ceneval identificará los programas-campus cuyo valor pueda clasificarse en alguno de los dos Estándares de Rendimiento Académico establecidos en el Padrón (Ceneval, 2013):

- *Estándar 1* (IDAP>1.8): Se ubican los programas-campus con, aproximadamente, 80% o más de sus sustentantes con TDS o TDSS.
- *Estándar 2* (1<IDAP<1.8): Se ubican los programas-campus con, aproximadamente, 60% o más de sus sustentantes con TDS o TDSS.

Para que los programas educativos de las IES puedan participar requieren que todos sus egresados, o bien, una muestra representativa de ellos (seleccionada de forma aleatoria por el Ceneval) presente el EGEL durante el periodo que comprende cada convocatoria. Por ende, aquellos programas que no cumplan con este requisito no podrán participar.

El Padrón fue una medida implementada para incrementar el uso del EGEL dentro de las IES bajo la idea de que el instrumento estaba pensado originalmente para retroalimentar la calidad de los programas educativos y, como tal, debía ser presentado por una proporción considerable de los egresados de cada programa educativo que hace uso de esta herramienta. Sin embargo, si bien el EGEL se utilizaba como medio regulador del egreso y la titulación, el Padrón reafirmó este uso y además generó su utilización con fines de propaganda y acreditación. Muestra de ello son las notas generadas por instituciones como: Instituto Tecnológico y de Estudios Superiores de Monterrey (2013), Universidad Autónoma del Estado de México (2013), Universidad Autónoma del Estado de Hidalgo (2014) o la Universidad Autónoma de Baja California (2013); dentro de las cuales se informa a la sociedad de los buenos resultados de los programas educativos al ser parte de este Padrón.

La creación del Padrón mediante el uso del IDAP pretende entonces aumentar el uso del EGEL dentro de las IES, en teoría, para retomar el esfuerzo inicial de esta evaluación como elemento de retroalimentación en busca de la calidad académica, sin embargo, su uso por parte de las IES retoma una postura propagandista, además de que el establecimiento de dos estándares para decidir si los programas educativos forman parte del padrón no parece aportar información que facilite la retroalimentación de los programas educativos. En lugar de ello el Padrón promueve la idea de que la calidad educativa se relaciona con la jerarquización o el ranqueo, que, como se comentó en capítulos previos, representa uno de los mayores problemas de la evaluación a gran escala. Adicionalmente conlleva a potencializar aún más su uso como medio administrativo, ya que la inclusión de los programas educativos en el Padrón (o incluso con solo la consideración de su aplicación) puede ser parte de la información para que los programas educativos sean

considerados como licenciaturas de calidad y se incorporen a programas como el PIFI, como previamente se mencionó; dicho de otra forma, el Padrón fomenta la utilización del EGEL como medio de acreditación y no como evaluación. Por tanto, dicha estrategia no se traduce necesariamente en la retroalimentación de los programas educativos, ya que la clasificación de la calidad de un programa educativo en un estándar determinado por el resultado de los estudiantes en una prueba no puede considerarse una *evaluación* en el sentido amplio del término.

A manera de cierre sobre este aspecto de análisis, debe tenerse presente que actualmente el uso del EGEL dentro de las IES se ha potencializado por lo que Salinas (2002) comentaba como la intención de hacer uso del avance (en términos acumulativos) de aprendizaje como elemento para la rendición de cuentas, ello debido a su funcionalidad como aspecto regulatorio del egreso frente a un proceso de masificación de la educación superior, así como debido a su utilidad como condicionante de obtención de recursos para las IES. Por ende, en la práctica el EGEL no necesariamente resulta ser una herramienta de análisis de la calidad, ya que el establecimiento de un ranqueo no es una herramienta que retroalimente las prácticas educativas. Bajo esta condición, vale recordar a Antúnez y Aranguren (2004), quienes consideran que dada la importancia de la evaluación, ésta debe centrarse en proponer mejoras educativas por medio de un proceso de reflexión de todos los elementos involucrados en el aprendizaje de los estudiantes. Este postulado adquiere importancia al tratarse de un elemento clave en el SES y uno de los ejercicios de evaluación a gran escala de mayor importancia en México. De esta forma, la tesis de Anderson (2005) que refiere a que los sistemas educativos basados en la rendición de cuentas se construyen sobre componentes que confluyen entre sí, como los objetivos, evaluaciones y recursos, comienza a verse como necesario para el avance del SES.

### **Concepción teórica del EGEL. ¿Cuál es el modelo que sustenta la actividad evaluativa del EGEL?**

Como se ha descrito, el EGEL tuvo un desarrollo altamente relacionado con las políticas educativas de los años noventa, así como debido a procesos sociales de masificación de la educación superior y a aspectos relacionados con los recursos y la rendición de cuentas. Si bien ello representa un eje importante que facilita la comprensión del contexto que motivó el surgimiento y que seguirá dando cabida al EGEL, es necesario reflexionar sobre el modelo teórico que lo sustenta y define la metodología de trabajo que ocupa.

Para comenzar a describir la perspectiva que fundamenta el EGEL es necesario iniciar con la definición de evaluación que esta prueba considera. Ceneval en su texto *Metodología Ceneval* (2013c; pág. 9) afirma que la institución se centra en “evaluar los conocimientos y habilidades adquiridos por los individuos en procesos de enseñanza-aprendizaje formales o no formales de los niveles educativos medio superior y superior”. La institución no muestra una definición de evaluación tal cual, sin embargo, el eje de su quehacer se centra en describir que realizará la *evaluación de conocimientos y habilidades*. Se puede inferir que se refiere a los estudiantes, aspirantes u egresados, dado que son las personas las únicas que pueden reflejar dichos constructos. Esta perspectiva coincide con sus postulados iniciales mencionados por Gago (1995) en lo relacionado a *realizar exámenes...para evaluar a los estudiantes que egresan...que solicitan ingreso...*. Desde este primer momento el Ceneval se define a sí mismo (y, por ende a sus proyectos) como una instancia de medición o técnica; lo cual representa un problema, ya que, como se ha comentado, reducir a la evaluación a un proceso de medición resulta limitado. De manera precisa, en los EGEL, se *evalúan el nivel de conocimientos y habilidades académicas de los recién egresados de carrera universitaria, permiten identificar si los egresados de licenciatura cuentan con los conocimientos y habilidades necesarios para iniciarse eficazmente en el ejercicio profesional*; ello de acuerdo con el Manual Técnico de los EGEL (Ceneval, 2013c). Claramente el objetivo del EGEL mantiene la perspectiva definida para todo el Ceneval, centrándose particularmente en los conocimientos y habilidades de los egresados de nivel licenciatura.

Para identificar con mayor precisión la postura de evaluación que subyace la actividad del Ceneval, pero en particular, del EGEL, baste con mencionar la amplia gama de definiciones sobre evaluación que la literatura puede ofrecernos para poder valorar la perspectiva teórica que fundamenta el quehacer del EGEL. De acuerdo con la revisión de literatura realizada por Ramos, Perales y Pérez (2009), existe una gran diversidad de definiciones de evaluación, cuyas diferencias no se dan por el objeto (si es de aprendizaje, de programas, etc.), a continuación se mencionan las citadas por dichos autores.

**Tabla 4.1 Definiciones de evaluación**

<b>Autor</b>	<b>Año</b>	<b>Definición</b>
Tyler	1950	Ver hasta qué punto se han alcanzado los objetivos.
Cronbach	1963	Proceso de recopilación y utilización de la información para tomar decisiones.
Scriven	1967	Proceso para determinar el valor de algo.
Stufflebeam	1971	Proceso de diseñar, obtener y proporcionar información

Tuckman	1975	útil para juzgar decisiones alternativas. Proceso mediante el cual se examinan las partes, procesos o resultados de un programa para ver si son satisfactorios respecto a los objetivos establecidos, a nuestras expectativas o a nuestros estándares de excelencia
Popham	1980	Consiste en emitir un juicio formal de los fenómenos educativos; es sistemática porque se diferencia de la evaluación informal que uno hace todos los días.
De la Orden	1982	Consiste en recoger y analizar información de una faceta de la realidad educativa para hacer un juicio a un patrón o criterio previamente establecido
Joint Committee	1981	Es el enjuiciamiento sistemático de la valía de algo.
Zúñiga	1994	Significa tomar una decisión sobre la orientación y el desarrollo de una acción a partir de criterios explícitos y consensuados.
Jiménez	1997	Es un proceso de recolección de información cualitativa y cuantitativa a través de instrumentos comparada con criterios establecidos que nos permite emitir juicios de valor.
Casanova	1999	Es la recogida sistemática y rigurosa de información con el juicio de emitir un juicio de valor respecto a ella. Esa valoración permitirá tomar decisiones consecuentes con el fin de mejorar.
Mateo	2000	Es la reflexión sistemática orientado a la mejora de las acciones de sujetos, de intervenciones profesionales, del funcionamiento institucional.
Juste	2000	Valoración a partir de criterios pre especificados de la información técnicamente diseñada y organizada, sobre cuántos factores relevantes integran los procesos educativos para facilitar la toma de decisiones de mejora.
Elola y Toranzos	2000	Es un esfuerzo de aproximación sistemática al objeto de evaluación. La información debe de tener un carácter retroalimentador, es decir, es un incremento progresivo en el conocimiento sobre el objeto evaluado.
Jornet	2001	Es parte del proceso de enseñanza-aprendizaje utilizado en la enseñanza por parte del profesorado, por ende, es fundamental en metodología didáctica.
Música	2003	Proceso de identificación y recogida de información relevante que puede ser cualitativa, cuantitativa, de manera sistemática, planificada, rigurosa, dirigida, objetiva, creíble, fiable y válida para emitir juicios de valor basados en criterios y referencias preestablecidos para determinar el valor del objeto educativo.
Castillo y Cabrerizo	2004	Es una característica inherente a toda actividad humana "intencional" y que requiere de objetividad y sistematización.
Expósito y	2006	Enjuiciamiento sistemático del mérito o valía de algo con

Olmedo Stake	2006	respecto a ciertos estándares comúnmente aceptados Proceso consistente de hallar pruebas de calidad para luego informar de esas pruebas a otras personas.
-----------------	------	---

Fuente: Elaboración propia con datos de Ramos, Perales y Pérez (2009).

Si bien pueden existir mucho más definiciones y más actualizadas, el ejercicio mostrado por Ramos, Perales y Pérez (2009) permite comprender la diversidad de perspectivas sobre la evaluación en educación, pero sobre todo, tener presentes los aspectos definitorios de éstas y cómo se relacionan con el EGEL. Las definiciones muestran algunos aspectos entre las cuales hay cierta coincidencia, como el énfasis en la observación o recopilación de información, el enjuiciamiento (o valoración en función de un estándar), la toma de decisiones y en los últimos años la retroalimentación. Cada uno de estos aspectos en los que coinciden diversas definiciones se resaltan de acuerdo con el momento en el cual se desarrollaron, siguiendo una aparente maduración del concepto de evaluación, desde la estipulación de analizar hasta qué punto se han alcanzado los objetivos en el caso de Tyler en 1950, hasta el énfasis en retroalimentación y búsqueda de la calidad de Elola y Toranzos en el 2000 y Stake en el 2006.

Dentro del objetivo del EGEL se puede apreciar su interés de evaluar (o valorar) la adquisición de conocimientos y habilidades de los sustentantes. Dicho objetivo implica la observación o recolección de información, además que la intención de identificar si los egresados *...cuentan con los conocimientos y habilidades necesarios para iniciarse en el ejercicio profesional*, lo que implica también la determinación de estándares que nos orienten sobre los conocimientos *necesarios* para el ejercicio profesional en nuestro país. Bajo este análisis, se puede observar que el EGEL tiene un objetivo muy similar a la perspectiva de Tyler, Tuckman, Popham, De la Orden, Joint Committee, Jiménez, Múgica y con Expósito y Olmedo. El objetivo del EGEL entonces, si bien se centra fuertemente en elementos de observación, también considera la de definición de estándares, abarcando así una perspectiva iniciada en 1950 hasta finales de los años 90. Pareciera entonces que el EGEL no considera dentro de sus objetivos ser parte del proceso educativo o el valioso elemento retroalimentador orientado hacia la calidad que justificó su origen.

Retomando nuevamente el Manual Técnico de los EGEL (Ceneval, 2013c; pág. 13), se comenta lo siguiente:

“El EGEL es una prueba nacional especializada por carrera profesional, cuyo propósito es identificar la medida en que los egresados de la licenciatura cuentan con los conocimientos y habilidades esenciales para el inicio del ejercicio profesional; de esta

forma, se asume que la información resultante de su uso sistemático contribuye a establecer distintos aspectos de gran importancia, como: a) el grado de idoneidad de cada egresado con respecto a un estándar de formación nacional; b) el nivel de eficacia de los programas y modalidades de formación profesional que administran las IES y c) indicadores de rendimiento a partir de los cuales, los organismos evaluadores y acreditadores respectivos, pueden identificar los méritos de cada programa de licenciatura de las IES.”

Frente a esta descripción se puede corroborar que el EGEL está enfocado en la observación y la comparación de lo observado con un estándar nacional. Claramente también se valora el resultado de los estudiantes como el indicador de mayor fuerza para hacer un juicio sobre la calidad educativa (entendida en este caso como la idoneidad de los egresados, la eficacia de los programas –en función de los resultados en una prueba– y los indicadores de rendimiento). Esto permite valorar la fuerte apuesta del EGEL por ser un elemento técnico enfocado en la medición y generación de indicadores cuantitativos; ello se puede observar desde el nombre de la evaluación, la cual se piensa solo como un *examen*. Considerando esta perspectiva de la evaluación, sobre todo debido al grado de importancia que esta evaluación representa para las IES mexicanas, las reflexiones de Antúnez y Aranguren (1998) resultan valiosas, en el sentido que los autores afirman que la evaluación del aprendizaje sigue una línea evidentemente positivista, desde los postulados de Tyler y la búsqueda del cumplimiento de objetivos, la evaluación del aprendizaje se ha centrado en la búsqueda de la existencia del conocimiento en los alumnos. A pesar de ello, la evaluación se ha enfocado en elementos puramente técnicos que no guardan relación con los paradigmas educativos que implican una dinámica entre alumnos y profesores (así como su contexto social). Frente a ello, Luria, Leontiev y Vigotsky (1973. p.15, en Antúnez y Aranguren, 1998) afirman que *...la actitud psicométrica se preocupa sólo del resultado (performance) de la prueba presentada en condiciones estandarizadas, y por lo tanto prescinde, de modo casi total, del proceso de construcción de la propia solución, basándose en el principio según el cual el resultado final es una “representación” fiel de todo el proceso de solución*. Es bajo esta mirada que la evaluación que se ha desarrollado por medio del EGEL tiende, como comentan Antúnez y Aranguren (2004), a una perspectiva que refleja el cumplimiento de objetivos de aprendizaje cómo la forma de develación de la realidad educativa de una persona; adicionalmente, como comentaba Salinas (2002), solo la medición del rendimiento de estudiantes no puede considerarse como evaluación.

A partir del establecimiento de su objetivo, el EGEL se describe a sí mismo como una *prueba*, y dentro de su Manual Técnico se incluye una descripción detallada de los elementos que le dan

validez y confiabilidad. Si bien más adelante se analizarán a detalle sus aspectos técnicos y metodológicos, es menester destacar su coincidencia con el enfoque tyleriano de evaluación. Como se comentó en el capítulo 1, existen diversos enfoques de evaluación que determinan una metodología en particular. La siguiente tabla presenta los enfoques y sus autores revisados a manera de recordatorio.

**Tabla 4.2 Modelos de evaluación y sus principales autores.**

<b>Modelo</b>	<b>Autor (es)</b>
Modelo centrado en objetivos	Tyler
Modelo científico	Suchman
Modelo CIPP	Stufflebeam
Evaluación iluminativa	Parlett y Hamilton
Evaluación centrada en el consumidor	Scriven
Evaluación comprensiva	Stake

Fuente: elaboración propia.

El EGEL como prueba, más que como una evaluación, pretende observar la congruencia entre el trabajo o aprovechamiento de los egresados de una carrera profesional y un estándar en particular. Retomando sus características principales descritas al inicio de este capítulo, el EGEL, al considerarse como *sensible a la instrucción*, supone que el aprendizaje de los estudiantes es una consecuencia resultante de un programa educativo, justo como lo percibe la perspectiva tyleriana (Salinas, 2002). Por otra parte, se debe considerar que la prueba EGEL sigue la tendencia clásica de procurar la mayor objetividad posible, es por ello que hace uso primordial de los reactivos *de opción múltiple*; la consideración este tipo de reactivos en sí mismo no implica una relación estrecha con la perspectiva del modelo de Tyler, su relación consiste en la reductibilidad o atomización de los conocimientos que se pretenden medir de cada egresado como insumo principal de la generación de reactivos. Si bien es un elemento que se describirá a detalle más adelante, baste por ahora relacionar la intención de reducir la complejidad del conocimiento del ejercicio de una profesión en instrucciones precisas para el desarrollo de reactivos de opción múltiple con la perspectiva tyleriana y su visión que afirma que el conocimiento se establece desde lo concreto (para la facilidad en su medición), así como la suposición de que el conocimiento está libre de valores y su reduccionismo es factible para fines de verificación empírica, que, por ende, resultará predecible en términos estadísticos (Giroux, 1990, en Salinas,

2002). Por supuesto, se debe considerar que la simplificación de las formas de evaluación tendrá como consecuencia la simplificación de la forma en que se percibe al aprendizaje, y es que bajo esta perspectiva los significados de una persona o de un profesional pueden ser operacionalizados de forma invariable, además que se pretende cuantificar la realidad con el fin de crear un conocimiento altamente formalizado (Popkewitz, 1988, en Salinas, 2002).

Por su parte, una de las características más reconocidas del EGEL es que es una prueba *estandarizada*, manteniendo una *esencia fija* (Ceneval, 2013b). La estandarización implica una sistematicidad en su desarrollo, construcción, aplicación y calificación, tal como se desarrolla en el modelo tyleriano, el cual valora la objetividad que la sistematicidad ofrece como la pretensión de una comprensión de una realidad ordenada (Salinas, 2002). Ahora bien, el EGEL en la actualidad es *criterial*, ya que en cada aplicación compara los resultados de cada evaluado con un estándar previamente establecido; ello facilita la objetividad característica de los instrumentos de medición y asegura que el juicio de los evaluadores sea relativamente neutral, sin embargo, bajo esta lógica los juicios entonces pueden desarrollarse solo con base en respuestas dadas a formatos cerrados.

Como se observa, el EGEL tiene una alta coincidencia con el modelo tyleriano; y si bien Tyler pretendía que la evaluación tuviera un impacto importante en la retroalimentación en los programas educativos y fue él quien estableció la principal diferencia entre medición y evaluación, considerando a la segunda como una acción enfocada a la retroalimentación y mejora, Ceneval sigue concibiendo a su evaluación del egreso de la licenciatura con fines de mejora de la calidad de la educación superior mexicana como sinónimo de examen o prueba, es decir, como una acción de medición de conocimientos y habilidades reductibles a su mínima expresión, alejándose a sí mismo del modelo de evaluación más básico con el cual tienen mayor relación. Si bien en los años 90, cuando iniciaron los EGCP, la posibilidad de realizar una evaluación en un contexto ajeno a la cultura evaluativa fue una opción viable que favoreció la implementación de la cultura de rendición de cuentas dentro de las IES, no se ha alcanzado el objetivo primario que le dio vida a este proyecto: “informar a la sociedad mexicana acerca de la calidad en la formación académica de los nuevos profesionistas, contribuir a la evaluación de la calidad de la educación superior, aportar información que permita definir acciones eficaces para mejorarla, ampliar el conocimiento del que disponen las instancias gubernamentales para establecer políticas de apoyo y asignación de recursos y fundamentalmente informar al propio sustentante acerca del nivel de formación alcanzado...” (ANUIES, 1993; pág. 1).

## Metodología para su elaboración

Habiendo analizado el contexto bajo el cual se ha desarrollado y mantenido el EGEL como una opción para las IES mexicanas, así como los postulados teóricos con los cuales guarda relación, es preciso abordar sus características metodológicas, cuya descripción permitirá una mejor comprensión de su desarrollo.

*Definición teórica del constructo de medida. ¿Cuál es la metodología que permite la definición teórica del constructo de medición?*

Habiendo reflexionado ya sobre cómo esta evaluación se concibe a sí misma como un instrumento de medición, se debe resaltar la importancia que Ceneval da a la definición de su constructo. Dentro de los procesos psicométricos o de medición de rasgos psicológicos, el constructo representa el eje de los instrumentos, y es que constituye la definición o construcción teórica de la serie y el conjunto de contenidos sobre los cuales se realizará una observación mediante instrumentos de medida (ya sea exámenes, cuestionarios, listas de verificación, etc.).

De acuerdo con Camacho y Sánchez (1997; pág. 145), los constructos normalmente son “abstracciones teóricas (estados mentales no directamente observables)”, que para que puedan ser medidos y valorados es menester precisar su contenido. La definición de un constructo puede hacerse a través de dos procedimientos no excluyentes: la aproximación inductiva, en la cual redactan muchos reactivos de un supuesto constructo y una vez que se contestan el investigador los agrupa de acuerdo a patrones de relación de respuesta, para con ello asignar categorías de constructo; y la aproximación deductiva, la cual desarrolla los reactivos con base en un elemento teórico que los sustente; en el caso del EGEL se hace uso de una aproximación deductiva. Adicionalmente, lo que la literatura indica es que en la definición de los constructos de medición se pueden considerar diversas fuentes de información tales como la revisión de literatura, el análisis de contenido (consiste en hacer preguntas a expertos sobre el tema y categorizar sus respuestas en dominios del constructo), el uso del juicio de expertos, el análisis de incidentes críticos (se pide a expertos que indiquen los comportamientos que muestran las personas en los extremos de un continuo de desempeño que puede representar al constructo), o bien, la observación directa (Camacho y Sánchez, 1997). Ceneval se enfrenta a un problema metodológico serio en el momento de la definición de su constructo, ya que en la actualidad al ser una prueba no curricular, no hace uso de los diversos planes y programas de estudio de las diversas

licenciaturas en las distintas IES a las cuales se dirigirá, sin embargo, recurre a técnicas útiles como el juicio de expertos y la observación directa, como se describe en este apartado.

De forma transversal a todo el proceso de elaboración del EGEL participan dos órganos muy importantes, encargados de dar validez disciplinar a los trabajos desarrollados: el Consejo Técnico (CT) y los Comités Académicos (CA), ambos bajo la orientación metodológica del personal del Ceneval. Hay un CT y un grupo de CA para cada EGEL. El CT se integra por entre siete y 25 representantes de las IES, colegios de profesionales e instancias empleadoras relacionadas con la disciplina nombrados por los rectores, presidentes o directores de cada institución previa invitación por parte de la Dirección General del Ceneval (Ceneval, 2012); mientras que los CA se encargan de la elaboración de las tareas especializadas relacionadas propiamente con la delimitación del contenido de los instrumentos. En particular, a los CT les compete delimitar el objeto de medida del instrumento (o constructo), definir el perfil del sustentante (que regularmente se define *ab initio* como los egresados de nivel licenciatura, o bien, estudiantes cercanos al egreso bajo la autorización de la IES para ello), determinar la estructura de la prueba y su organización, fijar los puntos de corte y establecer los niveles de desempeño para cada área de la prueba (con orientación técnica del Ceneval), así como determinar los criterios con los que se asignarán los Testimonios de Desempeño Satisfactorio y Sobresaliente (Ceneval, 2013). El CT se conforma procurando mantener un equilibrio entre instituciones y organismos públicos y privados y, en la medida de lo posible, también de su zona geográfica, ello para fines de cuidar la diversidad y representatividad de las diferencias estatales y culturales de cada institución participante. En el caso de los CA, éstos son expertos que laboran dentro de las IES que cuya responsabilidad es precisar contenidos que serán medidos en la prueba a través de la elaboración y validación de especificaciones, así como elaborar y validar los reactivos que pretenden incorporarse al EGEL (Ceneval, 2013). Regularmente los miembros de los CA son sugeridos por el CT, o bien, en su gran mayoría se incorporan al CA a través de su participación en talleres para el desarrollo de reactivos que se imparten dentro de las IES previo acuerdo entre Ceneval y las instituciones para fines de alimentar el número de reactivos en el banco del EGEL de interés, con los cuales se ensamblan las versiones de las pruebas para cada aplicación.

Dentro de los primeros pasos para la elaboración de un EGEL se encuentra la conformación del CT. Una vez que se cuenta con dicho grupo de expertos, el CT trabaja junto con el personal del Ceneval para la delimitación del objeto de medida (o delimitación del constructo). Todos los EGEL

se organizan en cuatro niveles de desagregación del conocimiento: Áreas, subáreas, temas y contenidos específicos (Ceneval, 2013). Para su definición se pregunta al CT sobre la organización de la práctica profesional; es decir, se espera que las áreas del EGEL correspondan a las principales funciones de la profesión o disciplina en cuestión. Por ejemplo, el EGEL-Psicología no se divide en las áreas temáticas tradicionales que se pueden localizar en los diversos planes de estudio, como: psicología experimental, psicología educativa, psicología clínica, entre otras; sino que fragmenta el conocimiento por medir en funciones del ejercicio profesional, a saber: Evaluación psicológica, Intervención psicológica e Investigación y medición psicológica. Existen otros ejemplos de EGEL cuyas disciplinas tienen una esencia más pragmática que la psicología que permiten valorar de mejor manera la lógica de la organización de contenidos del EGEL; por ejemplo, el EGEL-Turismo no se organiza en áreas temáticas tradicionales como: Agencias de viajes, Hotelería, Sociología del Turismo o Metodología de la investigación. En su lugar, el EGEL se organiza en: Desarrollo de productos y servicios turísticos, Gestión de organizaciones turísticas y Comercialización de productos y servicios turísticos. Como puede valorarse en estos ejemplos, se pretende que cada EGEL tenga una organización de contenidos de forma transversal, es decir, que no se aterrice en una temática o postura teórica en particular, de tal suerte que sus reactivos puedan ser desarrollados bajo una lógica centrada en la solución de problemas (como lo marcan las características del EGEL descritas al inicio de este capítulo) y no bajo un eje teórico. Si bien esta ha sido una forma en la que Ceneval ha procurado la unificación de la heterogeneidad de formaciones en educación superior existente en el SES mexicano, la constitución de las grandes áreas de la prueba se realiza bajo una metodología que no permite el cumplimiento de dicho objetivo, como más adelante se explica.

A pesar de ello, este primer paso de desarrollo del EGEL mediante el establecimiento de las grandes áreas o funciones coincide con la metodología general psicométrica. Por ejemplo, Pérez y Tornibeni (2004), mencionan que la elaboración del dominio de una prueba requiere un análisis conceptual del dominio o rasgo que se pretende medir; por ende, es necesario obtener o construir definiciones conceptuales del rasgo de interés y seleccionar los indicadores operacionales que permiten describirlo. En el caso del EGEL esta responsabilidad recae fuertemente en el CT en primera instancia, ya que son ellos, quienes con una visión general de la disciplina, pueden contribuir a la definición de las grandes áreas de ejercicio profesional. Una vez definidas las *áreas*, que son las funciones o ámbitos de actuación profesional, el CT trabaja la definición de *subáreas*, las cuales corresponden a las actividades que típicamente realiza el profesionista en cada una de

las áreas previamente definidas; una vez concluida esta labor, es tarea del CA la definición de los *temas*, que refieren a las tareas o acciones necesarias para llevar a cabo las actividades definidas en las subáreas, estos temas son validados por el CT. Es decir, la definición del constructo de medición o de los contenidos que incluirá el instrumento, constituye en una decantación del conocimiento, parcelado en función de las áreas inicialmente definidas. Previo al desarrollo del último nivel de atomización del ejercicio disciplinar, se desarrolla una encuesta para la validación social de los perfiles profesionales de los EGEL (Ceneval, 2009).

La encuesta para la validación social pretende ser una manera de homogeneizar las diferentes perspectivas de las formaciones profesionales heterogéneas que caracterizan a nuestro SES; en particular consiste en preguntar mediante un instrumento en línea a la comunidad de los egresados de todo el país de cada EGEL en desarrollo los siguientes aspectos sobre cada uno de los *temas* propuestos por el CA y el CT: a) *Importancia* para desempeñarse con eficacia a lo largo de su vida profesional; b) *Frecuencia* en la que realizaba cada tarea profesional a lo largo de un año; y c) *Estudio durante la licenciatura*. Mediante la combinación de dichos aspectos ponderados en función del tiempo del egreso del encuestado, se genera un indicador que permite al CT decidir sobre la inclusión o no de todas las tareas profesionales definidas y enjuiciadas por los egresados. Esta metodología le permite a Ceneval mantener una relativa validez social sobre sus instrumentos, y es que es relativa en el sentido de que no hay una muestra definida para ello, sino que la participación queda en función de la difusión que cada IES haga sobre sus egresados, así como de la voluntad de los mismos de aportar su experiencia para la conformación del objeto de medida del EGEL. De acuerdo con la experiencia que obtuve en la validación social de tres EGEL, una alta proporción de los participantes en la encuesta corresponde a los propios docentes de las IES, o bien, a personas que se encuentran laborando en el medio de la disciplina, pero que no necesariamente estudiaron el programa de interés (por ejemplo, contadores o ingenieros en alimentos que participaban en las encuestas del EGEL-Gastronomía), que son las personas a las cuales los programas educativos pueden acercarse de forma más sencilla. Si bien estos últimos son depurados del listado de participantes, esa acción reduce aún más la incipiente participación de los egresados. A pesar de ello, es a través de este procedimiento que Ceneval pretende la combinación de la técnica del juicio de expertos (por parte de los miembros del CT y del CA) y de la observación directa (simulada por medio de una encuesta en línea) para la definición de su constructo de medición.

No cabe duda que estos primeros pasos para la definición del constructo resultan innovadores y atractivos, sin embargo, no se puede asegurar que un estándar nacional, como pretende ser el EGEL, se base en la opinión de un conjunto de representantes institucionales, o bien, a partir de una encuesta cuya participación suele ser baja por parte de la población objetivo del instrumento: recién egresados de la licenciatura. Si bien el juicio profesional de los miembros del CT y del CA resulta de gran valía dentro y fuera del aula para la formación de futuros profesionistas, su opinión no asegura la representatividad del instrumento frente a la enorme diversidad de planes de estudio y posturas pedagógicas y didácticas que se encuentran dentro de cada IES, así como entre ellas; sobre todo si se considera la multiplicidad (y por ende, complejidad) de modalidades educativas que conforman el SES mexicano, como ya se explicaba en los párrafos introductorios de este trabajo. El problema se acentúa al considerar la diversidad cultural característica de nuestro país; en particular en disciplinas fuertemente influenciadas por el contexto económico y social, tal es el caso del Turismo, Gastronomía, Comercio Internacional, Derecho, Pedagogía, Ciencias de la Comunicación, Ciencia Política y Administración Pública, Relaciones Internacionales y Trabajo Social. Adicionalmente se debe considerar que hay disciplinas caracterizadas por diferencias entre perspectivas teóricas que definen metodologías de trabajo profesionales, como el caso de Psicología, Pedagogía, Trabajo Social, Relaciones Internacionales, Mercadotecnia, Gastronomía, Economía, Ciencias de la Comunicación o Ciencia Política y Administración Pública. En cualquier caso, cada EGEL pretende la definición de un mismo instrumento para una diversidad de conocimientos y métodos mediante un método que no asegura la cobertura de una enorme diversidad (cultural y didáctica), o bien, que se desarrolla sin la consideración de la misma. Adicional a esta problemática, la falta de coordinación de formaciones profesionales que caracteriza al SES mexicano representa una dificultad importante para la búsqueda de unificación con fines de evaluación.

No se debe olvidar que la evaluación es una actividad de mayor amplitud; como comentan Antúnez y Aranguren (2004), la evaluación es la reflexión crítica que involucra a una gran parte de los participantes en el proceso educativo sobre la teoría y la práctica del conocimiento con miras a la transformación. Es decir, es una acción inherente al proceso educativo, por ende, debe considerar entonces la realidad histórica-social-cultural, los sujetos actores, alumnos y docentes del proceso y el conocimiento interdisciplinario.

Como elemento final en el proceso de definición del constructo de medición del EGEL, una vez que se depuran los *temas* en función de los resultados de la encuesta para la validación social, el CA comienza la redacción de los *contenidos específicos*, los cuales se corresponden con los conocimientos y habilidades para realizar las tareas profesionales que se han contemplado. Estos contenidos se desarrollan con base en un formato de *especificaciones*, el cual es un documento que ayuda a que los CA operacionalicen los contenidos de la prueba y constituyen el insumo que será utilizado para una estrategia de elaboración masiva de reactivos.

*Elaboración de especificaciones. ¿Cómo es el proceso de elaboración de especificaciones por evaluar?*

Las especificaciones no son una propuesta teórica del EGEL. Es un término recurrente en el desarrollo de instrumentos de medición, en particular en aquellos enfocados a la educación. De manera cotidiana se utilizan dentro del Ceneval para el EGEL y otros proyectos que desarrollan. De acuerdo con el documento de la Metodología del Ceneval (2013b; pág. 12), las especificaciones de reactivos “ponen de manifiesto una definición operacional de conocimientos, habilidades o competencias que se pretende medir, así como un reactivo muestra”. Es decir, representan una descripción de cómo deben ser los reactivos del instrumento.

El desarrollo de estas instrucciones o *especificaciones* corresponde al CA. A partir de las *tareas* aceptadas por el CT, los miembros del CA bajo la tutela del personal del Ceneval enlistan los conocimientos y habilidades inmersos en cada tarea, y cada uno de ellos se expresa en términos de una *definición operacional*; es decir, cada conocimiento se redacta en términos observables para facilitar la generación de un reactivo de opción múltiple a través de esta definición. A continuación se presenta una especificación del EGEL-Turismo obtenida del Manual Técnico de los EGEL (Ceneval, 2013b).

**Figura 4.1 Ejemplo de especificación del EGEL en Turismo**

Clave de especificación	F3A3T2E1							
Función	Comercialización de productos y servicios turísticos							
Actividad	Aplicación de procesos para la comercialización							
Tarea	Determinación de la demanda del producto turístico de acuerdo con el análisis de mercado disponible							
Conocimiento y habilidad	Determinar las técnicas para pronosticar la demanda total de mercado							
Definición operacional	A partir de la descripción de la situación de mercado de un producto o servicio turístico, el sustentante determina la o las técnicas para pronosticar la demanda total							
<b>Clasificación</b>								
<b>Nivel de complejidad</b> (marque con "x")				<b>Tipo de contenido</b> (marque con "x")				<b>Reactivo resoluble por un sustentante con nivel de desempeño</b> (marque con "x")
Comprensión	Aplicación	Solución de problemas	Hechos	Conceptos	Principios	Procedimientos		
X						X	X	
<b>Reactivo muestra</b>								
Base	<p>Un hotel cuenta con los siguientes datos sobre el mercado nacional e internacional:</p> <ul style="list-style-type: none"> <li>• Existen 4,500,000 turistas internos y 12,800,000 turistas receptivos</li> <li>• Los turistas internos adquieren dos noches en promedio de hospedaje y los turistas receptivos siete noches en promedio</li> <li>• El costo por noche es de 57 USD para interno y de 48 USD para el receptivo</li> </ul> <p>Con el fin de pronosticar la demanda total del mercado nacional para el producto, es necesario utilizar las variables que determinan la demanda</p>							
Opción A								
Argumentación de la opción A	Incorrecta. En esta técnica se utilizan variables como aumento de la población o el ingreso per cápita, que es información con la que el hotel no cuenta.							
Opción B								
Argumentación de la opción B	Correcta. Requiere los datos de los compradores y sus compras para determinar la estructura del mercado al cual se dirige el hotel.							
Opción C								
Argumentación de la opción C	Incorrecta. Implica el uso de estimaciones lineales, que no es el insumo bajo el cual el hotel está determinando su demanda.							
Opción D								
Argumentación de la opción D	Incorrecta. Se identifican los factores del mercado correlacionados con el potencial del mercado, que son insumos que no está utilizando el hotel en este caso en particular.							
							<b>Respuesta correcta B</b>	
<b>Bibliografía sugerida</b>								
<ul style="list-style-type: none"> <li>• Kotler, P., et al. (2001). <i>Marketing</i>. México: Pearson.</li> <li>• Kotler, P. (2004). <i>Marketing para turismo</i>. México: Pearson.</li> </ul>								

Fuente: Ceneval (2013b)

Como se observa, la definición operacional pretende ser un fiel reflejo del conocimiento por evaluar (al cual, en este caso, le precede una decantación del área *Comercialización de productos y servicios turísticos*). Adicionalmente se describen elementos taxonómicos del conocimiento por evaluar, como el nivel teórico de complejidad y el tipo de contenido, con el fin que el elaborador de reactivos comprenda lo que se espera medir en cada reactivo que desarrolle. Se incluye un

ejemplo de reactivo, el cual pretende estar inmerso en un problema de índole profesional para la disciplina y que promueva la respuesta del examinado para así inferir la existencia de dicho conocimiento en la persona evaluada. Esta intención es tradicionalmente la perseguida por los psicómetras, en el sentido que se busca la observación objetiva de una muestra de comportamientos. Los estímulos o reactivos son una muestra del universo infinito de comportamientos del mismo tipo; a partir de las respuestas de la persona se pueden asignar valores numéricos que posteriormente serán comparados con los estándares para inferir que la persona cuenta con la aptitud que se pretende conocer (Cortada, 1999).

Existen dos tipos de pruebas clasificadas en función de su forma de calificación: con base en la norma y con base en criterio. En pruebas referidas a la norma, la prueba comienza su desarrollo con la elaboración de las tablas de especificaciones (tabla donde se relacionan los objetivos que se pretenden medir con los contenidos correspondientes), ello facilita la identificación del número de reactivos de la prueba, así como su redacción. En las pruebas referidas al criterio, como en el caso del EGEL, se define y delimita el dominio de conductas que corresponden a cada objetivo en particular; es decir, se enlistan aquellas tareas que la persona podrá realizar si el objetivo ha sido logrado (Pérez y Tornibeni, 2004).

El desarrollo de las especificaciones es un proceso crucial para el EGEL y es en donde establece una gran cantidad de recursos, ya que de éste depende el desarrollo de reactivos de forma sistemática y objetiva, ya que las especificaciones cumplen un rol de instrucciones para el desarrollo y la validación de los reactivos. Sin embargo, no se puede pretender que dar respuesta correcta a cada uno de ellos pueda traducirse en una inferencia directa respecto a que el egresado cuenta con el conocimiento o la habilidad que se pretende conocer. Frente a ello, Fredricksen (1984, en De la Cruz, Leyva y Seda, 2013) estableció que en una prueba, las respuestas correctas de los alumnos dependen de: (1) las coincidencias entre lo que el estudiante aprendió o recuerda (de todo lo que es posible aprender o recordar), (2) del currículo establecido, (3) de lo que los docentes enseñaron, y (4) de lo que se preguntó en la prueba. Bajo este argumento, el desarrollo de una evaluación a gran escala del egreso de la licenciatura no debe permanecer ajeno a la realidad social y cultural de los evaluados, las prácticas pedagógicas a las cuales se enfrentaron los egresados. La evaluación no solo se debe analizar los objetivos o efectos buscados, sino también los efectos secundarios y logros accidentales (Stake, 1975, en Stufflebeam y Shinkfield, 1993). Al respecto, Stake también afirma que los tests estandarizados “son a menudo inadecuados o

insuficientes para satisfacer los propósitos de una evaluación y deben frecuentemente ser sustituidos o completados con una variedad de métodos, incluyendo los soft y los subjetivos” (1975, en Stufflebeam y Shinkfield, 1993; pág. 238).

### **Elaboración de reactivos. ¿Cuál es el procedimiento que se sigue para el desarrollo de reactivos?**

La elaboración de reactivos de opción múltiple para el EGEL sigue un proceso fuertemente operativo, dado que la generación del marco de contenidos y todos los elementos teóricos de la prueba se han establecido previo a esta fase. Esta es una de las fases con mayor participación de docentes en los CA, sin embargo, ello no implica que se lleve a cabo un proceso comprensivo de desarrollo, ya que la aportación de los docentes va en el sentido de verter sus conocimientos y experiencias profesionales en los espacios disciplinares establecidos en las especificaciones previamente definidas y validadas por el CA y por el CT. Comienza entonces una tarea de producción masiva de reactivos con el fin de completar el número requerido por el CT en función de la cantidad de contenidos de la prueba, ello mediante talleres dentro de las IES con 16 horas de duración, divididas en dos sesiones de 8 horas cada una. En caso de que la prueba se aplique por primera vez se deberá desarrollar el triple de reactivos estimados para efectos de depuración psicométrica posterior a la aplicación. Frente a ello, el personal del Ceneval deberá llevar a cabo una labor de convencimiento con las IES para fomentar la participación de sus docentes como miembros del CA.

Existe una diversidad de formatos de reactivos dentro de cada EGEL, los cuales varían de acuerdo con su formato, a saber: *Completamiento*. Son enunciados o secuencias numéricas, gráficas o imágenes con elementos omitidos señalados con una línea, en donde el egresado debe seleccionar el elemento faltante de las opciones de respuesta. *Cuestionamiento directo*. Su base en un enunciado interrogativo o afirmación que el egresado debe completar mediante la selección de una respuesta. *Elección de elementos*. Consiste en la elección de algunos elementos de un listado de acuerdo con un criterio determinado; las opciones de respuesta se presentan como un subconjunto de los elementos enlistados. *Jerarquización u ordenamiento*. Un listado de elementos debe ordenarse con base en un criterio dado; las opciones de respuesta presentan los elementos en diferente orden. *Relación de columnas*. Se establecen dos columnas con elementos enlistados que el egresado debe vincular conforme a un criterio en particular; las opciones de respuesta

muestran posibles combinaciones de los elementos enlistados. Los reactivos se pueden presentar de forma independiente o como multirreactivos, éstos últimos implican un estímulo en común para una serie de reactivos asociados (Ceneval, 2013b). Independientemente del tipo de reactivo, siempre se desarrollan con cuatro opciones de respuesta donde solo una es la correcta; lamentablemente, en los documentos y manuales del Ceneval consultados no se hace explícita la argumentación del porqué de ello. Los lineamientos para el desarrollo de reactivos de los EGEL siguen las reglas tradicionales para la elaboración de reactivos que regularmente se expresan en la literatura sobre psicometría.

El proceso de elaboración de reactivos termina con su validación. Para dicho proceso, los reactivos desarrollados por el CA son enjuiciados por otros docentes de reactivos de IES distintas a las del elaborador. Cada reactivo se valora por tres miembros del CA con experiencia en desarrollo de reactivos. Este proceso se realiza como medio para llevar a cabo lo que en la literatura de medición se conoce como *validez aparente o face validity* (Nevo, 1985). Este proceso permite al Ceneval asegurar que el reactivo mide cabalmente la definición operacional de interés en función del juicio de los expertos disciplinares. Dentro de este proceso se procuran cuidar elementos de sesgo cultural con base en la opinión de los expertos; sin embargo, el objetivo principal es “acreditar la correspondencia plena del contenido del reactivo con lo establecido en la especificación” (Ceneval, 2013b; pág. 66); ello con fines de aseguramiento de estandarización dentro de sus procesos, con lo que se consolida el discurso de objetividad buscado por el Ceneval.

Una vez que se consigue el desarrollo de reactivos mediante estrategias masivas de producción de los mismos, el banco conformado representa entonces una muestra de los conocimientos y habilidades de una disciplina atomizados, los cuales, en conjunto, teóricamente representan el continuo del constructo definido en los inicios del desarrollo del EGEL. De esta manera, el Ceneval realizará inferencias de las respuestas de los egresados respecto a su dominio de las funciones profesionales establecidas por el CT y validadas por la comunidad de egresados. Por ende, en teoría los resultados de un egresado en el EGEL serán un reflejo de su nivel de eficiencia para poder hacer frente a una vida laboral, lo cual sería el indicador más fiable de la calidad educativa recibida. Este será un aspecto a revisar más a detalle al analizar las formas de establecimiento de estándares y de emisión de resultados.

## Proceso de aplicación. ¿Cómo se desarrolla el proceso de aplicación?

El proceso de aplicación del EGEL es una fase operativa que se implementa de forma cuidadosa, ya que de ella depende también que el proceso del EGEL se desarrolle de forma estandarizada. La logística de aplicación que se reporta en la Metodología Ceneval (2013c) incluye: *Planeación y programación*. Dentro de este proceso las IES registran en línea a sus sustentantes, se establecen costos y fechas de aplicación, así como sedes lo suficientemente grandes para albergar a los egresados que se registren. Existen cuatro fechas al año llamadas “nacionales”, dentro de las cuales se abren sedes estatales dentro de las IES con mayor capacidad e infraestructura. *Registro*. En esta etapa las IES registran a los sustentantes que presentarán la prueba. *Capacitación y asignación de auxiliares para la aplicación*. Dentro de la institución se selecciona a los auxiliares de aplicación (normalmente conocidos como “aplicadores”) que irán a las IES a administrar los EGEL; este proceso suele ser muy sencillo, en virtud que los aplicadores están familiarizados con los procedimientos estandarizados del EGEL y las precauciones de seguridad que implica una prueba de alto impacto, debido a que reciben capacitación en cada aplicación a la que son invitados. *Logística y preparación de materiales de aplicación en papel y en línea*. Este proceso implica la organización de materiales (sin personalización) en cajas en función de la demanda requerida en cada IES de acuerdo con los registros de cada una de ellas, así como su distribución entre aplicadores junto con los formatos de control de asistencia y de entrega-recepción de los cuadernillos. *Ejecución y control de la aplicación*. Los aplicadores administran las pruebas dentro de las instalaciones de las IES; la aplicación dentro de cada grupo puede incluir diversos EGEL, es decir, si la demanda así lo solicita, dentro de cada grupo se pueden aplicar los EGEL de tantas disciplinas como hayan sido requeridas, ello en virtud que los procedimientos de aplicación son los mismos en todos ellos, ya que su formato de presentación (ya sea en línea o en lápiz y papel) no representa diferencia alguna entre disciplinas. El número de sustentantes participantes en cada aplicación es muy variable, en función que puede ser desde un sustentante hasta cientos de ellos.

En general el proceso de aplicación es muy similar a lo que tradicionalmente se realiza en cualquier institución en la cual se apliquen instrumentos de evaluación del aprendizaje a gran escala. Tanto en Ceneval, como en otras instancias como el INEE, la UNAM o cualquier instancia educativa, los procesos de aplicación requieren la organización de la distribución de sustentantes en los espacios destinados, cuidar que las condiciones de las aulas no interfieran en la concentración de las personas (como aspectos de iluminación, ventilación y ruido), evitar el plagio

de materiales y llevar controles de asistencia de sustentantes (y evitar la suplantación de identidad). En particular se cuida que las instrucciones de aplicación sean las mismas entre cada grupo de aplicación, lo cual normalmente se hace a través de guiones que los aplicadores tienen que leer al pie de la letra, lo cual da cierta seguridad a las instancias evaluadoras respecto a que las condiciones de aplicación son las mismas para asegurar con ello una relativa validez del proceso de aplicación.

Los controles mencionados se establecen con la finalidad de buscar la estandarización de los procesos dentro y entre cada EGEL. Para asegurar el control de los estándares de aplicación Ceneval hace uso de diversos formatos de control que son manipulados directamente por una figura denominada *Coordinador de aplicación*, el cual es una persona responsable que va a cada instancia aplicadora a organizar y monitorear que los procesos se lleven a cabo de la forma esperada. Esto permite que el centro de evaluación pueda tener una relativa certeza que las condiciones de aplicación de todos los participantes son semejantes como para poder afirmar que el error estándar en la medición no sea una entera responsabilidad del proceso de aplicación; ello a pesar que siempre habrá diferencias en la aplicación en cada grupo, sin embargo, mediante el proceso descrito se busca la minimización de ese riesgo.

### **Proceso de definición de puntos de corte y calificación. ¿Qué metodología se utiliza para la definición de puntos de corte y emisión de calificación?**

Una vez que la aplicación se lleva a cabo, los aplicadores son responsables de regresar los materiales de aplicación a las instalaciones del Ceneval, donde se llevan a cabo dos procesos simultáneos. En primer lugar, para las pruebas a lápiz y papel, se leen las hojas de respuesta de cada persona de forma electrónica para conformar una base de datos con las respuestas seleccionadas en cada reactivo; por su parte, las respuestas de la aplicación en línea están ya de forma electrónica, lo cual facilita su tratamiento. En las aplicaciones en línea las IES deben contar con un aula con equipo de cómputo necesario para que se pueda ejecutar la prueba. Una vez que los estudiantes terminan de contestarla, las respuestas se guardan en automático y son enviadas por parte del aplicador al Ceneval para su tratamiento psicométrico.

Mientras los cuadernillos utilizados para la aplicación son almacenados durante un año para posibles consultas, la base de datos con las repuestas se envía a las áreas académicas del Ceneval para su análisis. Los datos generados tienen dos propósitos fundamentales: revisar la calidad

psicométrica de los reactivos y emitir la calificación de los sustentantes. A continuación se presenta una descripción de cada uno de ellos.

La revisión de la calidad psicométrica de los reactivos se realiza independientemente de la modalidad seleccionada (e lápiz y papel o en línea) y si bien es un proceso que se lleva a cabo en cada aplicación de cada EGEL, el análisis más importante es el resultante de la aplicación piloto, ya que de ello depende la conformación del instrumento. Como comentan Muñiz, Fidalgo, Cueto, Martínez y Moreno (2005), la calidad de la medición tiene dos indicadores principales: la validez y la confiabilidad; los métodos psicométricos nos permiten valorar la calidad de la prueba en función de estos dos elementos.

Todas las mediciones psicológicas se realizan con cierto grado de error, por lo que el conocimiento de la confiabilidad permite conocer la precisión con la cual se está desarrollando una medición. La confiabilidad o consistencia se puede afectar por la longitud del test, la variabilidad de la muestra e incluso el nivel de la persona en la variable medida. En el caso del EGEL, la consistencia interna de cada una de las áreas de la prueba se estima mediante el alfa de Cronbach (Ceneval, 2013b). Este indicador se utiliza en el ámbito de la psicometría desde los años cincuenta y permite estimar si la proporción de la variabilidad de la prueba se debe a factores en común de los reactivos (Cronbach, 1951), es decir, permite conocer la consistencia interna de la prueba, o bien, si los reactivos utilizados tienen una variabilidad similar, lo que asegura una confiabilidad. Para el caso del EGEL, el alpha de Cronbach aceptable para cada una de las áreas de cada examen es de 0.9 o más (en una escala en la cual el 1.00 significa una confiabilidad o consistencia de medida perfecta).

Por su parte, la validez refiere a que el instrumento mida lo que se pretende, es decir, que los puntajes que se obtengan reflejen fielmente el constructo de interés y no un constructo similar. Normalmente los elementos analizados en validez son: la validez de contenido, predictiva y de constructo. La primera de ellas refiere a la necesidad de comprobar que los reactivos realmente están representando el espectro del constructo evaluado; es decir, todos los atributos o indicadores que lo conforman. La validez predictiva refiere a la comprobación que el test predice cierto comportamiento, para ello se calcula la correlación entre el test y algún criterio o el comportamiento en cuestión. La validez de constructo refiere a la obtención de datos en la cual se compruebe que el constructo evaluado no es espurio; esto normalmente se realiza por medio de pruebas estadísticas, tales como el análisis factorial.

El análisis de los reactivos del EGEL se hace regularmente mediante la Teoría Clásica de los Tests (TCT). Lo que se pretende es conocer los parámetros cuyas características estén relacionadas con las propiedades y finalidad de la prueba. Los indicadores clásicos para determinar la calidad de un reactivo son los siguientes: *Dificultad*. Un reactivo será sencillo o difícil de acuerdo con la proporción de personas que lo contesten adecuadamente (en función, por supuesto, del número de personas que intenten contestarlo). El índice se construye de la siguiente manera:

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

A: Número de personas que aciertan el ítem  
N: Número de personas que intentaron responder el ítem.

La dificultad de una prueba debe mantener ciertos parámetros en función de su objetivo. Los reactivos que resultan en extremo sencillos o difíciles no son funcionales, ya que no permiten tener un adecuado nivel de discriminación del constructo. *Discriminación*. La idea de una prueba es escalar a las personas en función de su habilidad, por ende, un instrumento debe contener aquellos reactivos que permitan diferencias a las personas que cuentan con la habilidad requerida de aquellas que no. El índice de discriminación se obtiene de la siguiente manera:

$$D = p_+ - p_-$$

p<sub>+</sub>: Proporción de acertantes al ítem del grupo superior.  
p<sub>-</sub>: Proporción de acertantes del grupo inferior.

El valor oscila entre -1 y +1, mientras más alto sea el valor, mayor discriminación del reactivo y, por ende, mayor utilidad del mismo para afirmar que mide distinguiendo a los estudiantes que cuentan con el conocimiento medido de quienes no; es decir, la prueba se vuelve precisa. *Correlación*. Cuando las variables de los resultados de una prueba son continuas, pero las respuestas a un reactivo son dicotómicas se hace uso de la correlación biserial-puntual<sup>19</sup>, la cual permite valorar la relación entre cada reactivo y el resto de la prueba, es decir, si los patrones de respuesta entre la unidad y la totalidad son similares; su cálculo se hace mediante la siguiente fórmula:

---

<sup>19</sup> Llamada de esa forma debido a que permite relacionar una variable continua (el puntaje total de la prueba) y una discreta (la respuesta particular a una pregunta, la cual regularmente acepta solo dos valores: correcto o incorrecto).

$$r_{bp} = \frac{X_A - X_T}{S_x} \sqrt{\frac{p}{q}}$$

$X_A$ : Media en la prueba de las personas que acertaron el reactivo.

$X_T$ : Media total de la prueba.

$S_x$ : Desviación estándar de la prueba.

$p$ : Proporción de personas que acertaron al reactivo

$q$ : 1-p

Lo que la literatura indica es que mediante el cálculo de estos estadísticos es posible identificar la calidad de cada reactivo, ya que está en función de su nivel de dificultad, de discriminación del conocimiento de interés, así como de su relación con el conjunto de la prueba. En el caso del Ceneval, los psicométricos resultan ser el eje más relevante de análisis para la calidad de un reactivo.

Existe otro modelo llamado Teoría de Respuesta al Ítem (TRI), el cual no se basa en la suma de todas las respuestas de los estudiantes para la obtención de los estadísticos de discriminación y dificultad, sino que hace uso de la estimación de probabilidades de respuesta para el conocimiento de la habilidad de los estudiantes, las cuales no se representan como un puntaje o la media de un puntaje, sino como la probabilidad de dar una respuesta correcta a un reactivo en función del nivel de dificultad que el reactivo presente. El modelo estadístico que subyace a la TRI es extenso y complejo como para presentarse dentro de este trabajo, sin embargo, puede ser consultado de forma amplia en Muñiz (1997); para el objetivo de este trabajo baste mencionar que la lógica que define a esta metodología psicométrica se basa en el uso de probabilidades de respuesta con fines de estimación de habilidades de los evaluados, ello facilita el juicio sobre la habilidad de los estudiantes, ya que no están en función del total de personas que presentan la prueba (como sí sucede con la TCT), sino que las habilidades se obtienen mediante inferencias de probabilidades de respuesta a reactivos con dificultades similares a los que ha podido contestar de forma adecuada, sin considerar así al resto de la población para ello. Para el caso de los psicométricos (como dificultad o discriminación) igualmente se estiman mediante el cálculo de probabilidades de respuesta que los estudiantes dan, en este caso, necesariamente se consideran las respuestas de los estudiantes, no de forma directa, sino estimada mediante probabilidades de respuesta.

Cada que se realiza una aplicación de un EGEL se lleva a cabo un análisis psicométrico del comportamiento de los reactivos incluidos en calidad de nuevos (no aplicados previamente), con el fin de identificar su calidad de medida considerando los indicadores descritos. En este análisis se

considera la TCT, así como la TRI, sin embargo, en el trabajo cotidiano se hace un mayor uso de la TCT, ya que representa una forma menos compleja de interpretar el comportamiento de los reactivos. El análisis que se realiza de cada reactivo es un proceso particularmente atendido cuando se desarrolla un EGEL por primera vez, ya que de este análisis dependerá en gran medida la constitución de la estructura del EGEL.

El personal del Ceneval revisa las características psicométricas de cada reactivo con fines de descartar aquellos que tengan parámetros no deseables. Por ejemplo, una correlación negativa indicaría que el reactivo es confuso, ya que su patrón de respuesta no sigue el esperado; es decir, los estudiantes con bajos puntajes lo contestan correctamente, mientras los estudiantes con altos puntajes lo erran, lo cual es contrario a lo que se espera observar. Este tipo de casos son desechados y con ello comienza un proceso iterativo, en el que se analizan nuevamente los reactivos sin la consideración de aquellos que han sido desechados; este proceso es conocido como *Calibración psicométrica*. La calibración permite que los análisis (que como se observó en las fórmulas, se estiman en función de las respuestas de los evaluados a los reactivos) se desarrollen solo con reactivos con patrones de respuesta más apropiados; con ello se facilita que los reactivos que al final sean considerados para la prueba tengan los mejores parámetros posibles, ya que sus datos se estimarán con reactivos seleccionados y filtrados mediante un estándar en particular. El proceso de calibración de reactivos es una práctica cotidiana en la definición de instrumentos de medición psicológica (que, por supuesto, incluye a las pruebas que miden aprendizaje) ya que favorece la precisión del instrumento final. El problema con la calibración psicométrica desarrollada dentro del Ceneval es que no hay mucho espacio para la revisión cualitativa de los reactivos con parámetros no deseables, lo cual permitiría comprender el por qué se dan patrones de respuestas no esperados en reactivos que ha sido aprobados por un CA, sino que la función de la calibración consiste únicamente en el ajuste de los parámetros cuantitativos de los reactivos que se van filtrando en cada análisis. Si bien la calibración es necesaria en virtud a la dependencia que el desarrollo de instrumentos tiene con las respuestas de los sustentantes, se excluye el elemento pedagógico de revisión de los reactivos no ajustados al modelo psicométrico, dando así mayor relevancia a la función técnica de la medición *en sí misma* que al modelo teórico del constructo que se pretende medir.

Una vez que el personal del Ceneval ha depurado los reactivos considerando aquellos con mejor ajuste al modelo psicométrico, se realiza una reunión de definición de puntos de corte y niveles de

desempeño con el CT del EGEL en cuestión. Esta actividad, si bien no está teóricamente fundamentada en los textos técnicos publicados por el Ceneval, se debe desarrollar con el fin de contar con un referente respecto al rendimiento de un sustentante. Como comenta Stake (2006), los estándares son inherentes a la evaluación y la mayoría de las veces no son explícitos, por lo que deben acordarse, por lo que el desarrollo de una reunión para la definición de puntos de corte y niveles de desempeño permite el establecimiento del estándar contra el cual se compararán los puntajes obtenidos por cada egresado participante.

Epistemológicamente hablando, el establecimiento de un estándar representa la construcción de un modelo de análisis, de un marco de visión sobre una realidad educativa en particular, la cual funge como referencia respecto a desde qué perspectiva o realidad se está realizando un juicio sobre el desempeño de los sustentantes. Como comenta Salinas (2002), evaluar significa juzgar parcelas de realidad que se han encasillado; es decir, hacer una valoración con fines de retroalimentación a partir del establecimiento de una línea divisoria entre todo lo posiblemente observable que encasille nuestro juicio. Siendo así, en el EGEL el establecimiento de estándares permitirá fragmentar toda la posible realidad de *ser* de un egresado (definida ya previamente mediante el establecimiento de las Áreas, Subáreas y Temas, proceso que también excluye otras posibles formas de *ser* egresado) en niveles de desempeño. Mediante el proceso de definición del constructo y de los niveles de desempeño se establece un espacio de realidad que sirve como base de juicio, aunque, como dice Stake (en Stufflebeam y Shinkfield, 1993), en educación (y en general) hay diferentes realidades.

Bajo esta consideración epistemológica de los estándares y niveles de desempeño, son elementos necesarios para la interpretación de puntajes de una prueba; prácticamente en todas las pruebas con referencia al criterio se requiere del establecimiento de estándares; en particular, este tipo de definición de estándares son acordes para sistemas educativos que resaltan la rendición de cuentas (Anderson, 2005). Lo más común es tener entre tres y cuatro categorías de desempeño (por ejemplo, “debajo del básico”, “básico”, “adecuado” y “avanzado”), (Hambleton, Jaeger, Plake y Mills, 2000). En el caso del EGEL se clasifica el resultado de los estudiantes en tres niveles de desempeño: Aún no satisfactorio (ANS), Satisfactorio (DS) y Sobresaliente (DSS), aunque los documentos oficiales publicados por el Ceneval no dan una justificación del uso de tres niveles de desempeño. Dentro del proceso general en el desarrollo de pruebas, para la definición de cada nivel se establece un punto de corte que permite la fragmentación del continuo de los contenidos o

de las habilidades medidas mediante los reactivos calibrados. Para el establecimiento de puntos de corte existen diversos métodos; considerando los textos de Norcini (2003) y Hambleton, Jaeger, Plake y Mills (2000), los métodos más utilizados son los siguientes.

*Método del porcentaje ajustado.* En este método se solicita a cada juez que indique el porcentaje de sustentantes que pudieran “aprobar” el instrumento; los jueces comienzan una discusión con el fin de acordar los estimados. Una vez terminada la discusión las tasas de “aprobación” de los jueces se promedian para obtener un punto de corte. Este método es de fácil aplicación y fácil entendimiento por parte de los jueces; sin embargo, los niveles de desempeño que se utilicen serán independientes de los contenidos de la prueba.

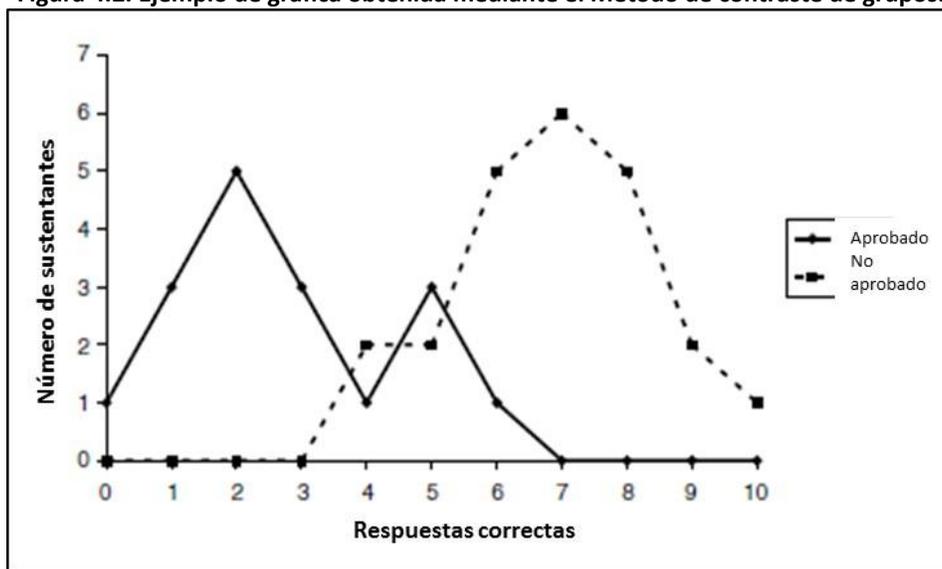
*Método Angoff.* En este método se pide a los jueces que definan las características del grupo limítrofe de sustentantes que divide cada nivel de desempeño. Posteriormente se considera la dificultad e importancia del primer reactivo de la prueba; cada juez estima el porcentaje de sustentantes limítrofes que responderán correctamente dicho reactivo. Los jueces discuten sobre esto mismo para cada uno de los reactivos incluidos en la prueba. Las estimaciones de los jueces se promedian para cada reactivo y los puntos de corte se obtienen considerando la suma de esos porcentajes. Este método es relativamente fácil de utilizar, sin embargo, no existe una base firme para las estimaciones que realizan los jueces, además que para pruebas largas este método puede resultar cansado.

*Método de Angoff extendido.* Este método se desarrolla para pruebas politómicas. Para fines psicométricos existen dos tipos de pruebas en función del formato de las respuestas de sus reactivos: las dicotómicas y las politómicas. Las dicotómicas (Domínguez, 2013) tienen un solo umbral que asume una ausencia total del conocimiento o una total presencia (normalmente representadas en reactivos de opción múltiple, donde una opción es la correcta y el resto asume una ausencia total del conocimiento de quién las selecciona). Por su parte, las politómicas infieren un umbral de respuesta, de tal suerte que cada opción de respuesta incluida asume un valor; un ejemplo de ello son las escalas tipo Likert o las rúbricas. En este método los jueces estiman un puntaje típico que un estudiante o sustentante limítrofe obtendría en una pregunta en particular; con base en ello se obtiene el porcentaje de los estimados de desempeño de los jueces. Los porcentajes de los jueces pueden sumarse para establecer un nivel de desempeño para la evaluación, o bien, pueden ser ponderadas estimando la importancia o valor de una pregunta individual. Su gran ventaja es su simplicidad de implementación, así como la posibilidad de

ponderar las preguntas de interés de los jueces; sin embargo, el método pudiera ser demasiado reduccionista, parcelando el conjunto de contenidos en partes independientes.

*Método de contraste de grupos.* Para poder utilizar este método inicialmente se debe seleccionar una muestra aleatoria de sustentantes. Con ello los jueces consideran las respuestas del primer sustentante en todas las preguntas de la prueba; en conjunto, los jueces determinan si el desempeño del sustentante es suficiente para “aprobar”. Este proceso se realiza para todos los sustentantes seleccionados en la muestra inicial. Posterior a que se realizan los juicios para todos los sustentantes seleccionados, se grafican todas las decisiones y con ese apoyo gráfico se determina los puntos de corte. Si bien es un método que requiere mucho tiempo para su implementación y que los juicios realizados pueden ser subjetivos, su implementación es de relativa simplicidad. La siguiente figura muestra un ejemplo sobre cómo es que este método genera una visión gráfica que facilita la determinación de los puntos de corte para los niveles de desempeño.

**Figura 4.2. Ejemplo de gráfica obtenida mediante el Método de contraste de grupos.**



Fuente: Norcini (2003)

Como se observa en la figura 4.2, se grafican los sustentantes que “aprueban” (línea punteada), así como los sustentantes o examinados que no lo hacen (línea continua). Con base en ello los jueces pueden llevar a cabo la determinación de puntos de corte que permitan diferenciar los niveles de desempeño en la prueba.

*Método Hofstee.* Este método confluye métodos relativos y absolutos; se solicita a los jueces que indiquen los puntos de corte mínimos y máximos aceptables, así como las tasas de “no aprobación” aceptables. Los resultados de cada juez se combinan para la obtención de los puntos de corte definitivos.

*Método de media estimada y distribución del puntaje esperado.* Este método es similar al de Angoff extendido, ya que requiere que los jueces estimen el número mínimo de puntaje para los estudiantes limítrofes, además que requiere que se estime la distribución de puntajes de los estudiantes para cada uno de los niveles de desempeño estipulados. La ventaja de este método radica en que se obtiene mucha información sobre el desempeño esperado en cada uno de los niveles de desempeño, mientras que su desventaja radica en que puede resultar algo complicado de entender por parte de los jueces.

*Método de mapeo de reactivos (Bookmark).* En este método se presenta a los jueces una escala del desempeño de los sustentantes; se asume que los estudiantes con mayor habilidad se desempeñarán mejor que aquellos con poca habilidad. Para hacer esta conclusión se requiere hacer uso de la TRI, con fines de estimar la habilidad de cada persona evaluada. El juicio de importancia radica en decidir el nivel esperado para cada nivel de desempeño; para ello se ordenan todos los reactivos de la prueba en función de su dificultad y, con base en ello, los jueces establecen los reactivos que representan una diferencia entre un nivel de desempeño y otro. En función de esa decisión, se delimitan niveles de desempeño bajo las habilidades que reflejan el nivel de dificultad de los reactivos seleccionados; en concreto, los reactivos marcadores elegidos por los jueces serán los que delimiten los límites de habilidad para definir los niveles de desempeño.

*Método de selección de trabajos.* Este método se usa para tareas o actividades que se codifican de forma politómica. Los jueces seleccionan los trabajos que representan de forma ilustrativa el desempeño de los sustentantes limítrofes en cada área evaluada. El porcentaje de los puntajes seleccionados como representativos de un desempeño limítrofe serán considerados como el valor mínimo para lograr cada uno de los niveles de desempeño.

*Método holístico.* En este método se solicita a los jueces que consideren el trabajo de los sustentantes y decidan aquellos que representan los mínimos de cada nivel de desempeño (como en el método de selección de trabajos), solo que, en este caso, se considera la totalidad de tareas

que puede desarrollar un sustentante para poder valorar en conjunto todos los posibles aspectos que den cuenta de su desempeño, evitando así fragmentar la valoración sobre ello.

*Método analítico.* Este método se utiliza para valorar aspectos que puedan ser calificados de forma politómica. A los jueces se les presenta un conjunto de respuestas que representan toda la posible variedad de su desempeño dentro de cada una de las secciones de la prueba; una vez que se hacen juicios de forma individual, se toman decisiones de forma colegiada analizando las causas de las posibles disparidades entre sus juicios. Con base en esta discusión se establecen estándares de desempeño para cada sección de la prueba que se suman para conocer el desempeño de los sustentantes en toda la prueba. Si bien es un método de fácil entendimiento, puede consumir mucho tiempo.

*Método del perfil dominante.* Este método puede utilizarse para pruebas con diversas áreas y que se puedan calificar de manera politómica. Mediante una estrategia de consenso, se pide a los panelistas que acuerden la política que determinará qué puntajes son los críticos para que los sustentantes “aprueben” la evaluación. Las decisiones individuales sobre los criterios de aprobación se revisan uno a uno y se identifican aquellas que tengan componentes que representen a la mayoría. Posteriormente se muestra a los jueces el impacto cuantitativo de sus decisiones. Si bien este método permite generar políticas o decisiones de estándares de forma colegiada, puede ser que la decisión final de necesariamente represente a todo el grupo de jueces, lo cual en algún momento pudiera ocasionar molestias entre los especialistas.

*Método del juicio político.* Para su desarrollo se capacita fuertemente a los jueces para que puedan valorar los tipos de puntaje de los sustentantes dentro de un ejercicio en particular, de tal forma que cuando se les presenta un tipo de puntaje los jueces puedan entonces formarse una imagen mental del desempeño general del sustentante. Cuando los jueces emiten su opinión ésta se cuantifica para generar modelos matemáticos (como la regresión) para estimar el desempeño de los sustentantes. Una vez que se ha hecho esto para todos los sustentantes, se presentan nuevamente los tipos de respuesta de forma ordenada, con el fin que los jueces determinen el puntaje mínimo esperado dentro de la prueba. Este método es flexible que da la oportunidad de dar importancia a algunos contenidos de la prueba en función de las opiniones de los jueces; por otra parte, puede ser un método que implica mucho tiempo, además que algunos jueces pudieran experimentar cierta frustración, ya que la decisión individual sobre los estándares recaerá más en elementos técnicos que en la justificación del porqué de las políticas que los orillan a esa decisión.

*Método de juicio directo.* Este método se utiliza para realizar juicios sobre un conjunto de tareas, aunque cabe la posibilidad de utilizar como insumo puntajes dicotómicos. Inicialmente se capacita a los jueces para que éstos comprendan el significado de los puntajes que los sustentantes pueden obtener en cada tarea o ejercicio. Los jueces determinan el peso que se asignará a cada uno de los ejercicios (lo que influirá en su cálculo final), así como el puntaje del desempeño general necesario para “aprobar” la evaluación.

Como se puede apreciar, existe una gran diversidad de métodos en la literatura, aunque muchos de ellos exigen una medición politómica; por supuesto, todos ellos tienen sus ventajas y desventajas, desde una fuerte dependencia a la población que presenta la prueba, como en el método de contraste de grupos, el *Bookmark*, el de selección de trabajos y el analítico. Dentro de estos métodos los estándares de desempeño dependerán prácticamente en su totalidad de las personas que presentan la prueba; aspecto que, curiosamente, una evaluación estandarizada pretende evitar. Por otra parte, hay métodos que tienen una gran dependencia con el juicio personal de los invitados a estos procesos, tal es el caso del método del porcentaje ajustado, el Angoff (y el Angoff extendido), el Hofstee, el de media estimada, el holístico, el de perfil dominante y, por supuesto, el del juicio político. En todos los casos la determinación de puntos de corte para definir niveles de desempeño conlleva ciertas implicaciones que se deberán tomar en cuenta al momento de decidir qué método utilizar.

En el caso del EGEL se hace uso del método *Bookmark*, aunque en sus documentos oficiales y cuadernos técnicos no se especifica el porqué de esta decisión. Como se comentó, el método *Bookmark* hace uso de dos elementos básicos: la dificultad de los reactivos (estimada a través de la TRI), así como el juicio de los expertos para definir el reactivo marcador que determina el conocimiento mínimo necesario en cada nivel desempeño. Si bien Ceneval no tiene publicado cómo se desarrolla el método, es muy similar a lo definido en la literatura. De acuerdo con Karantonis y Sireci (2006), el método implica la inclusión de los reactivos depurados en *booklets* o cuadernillos; los reactivos son ordenados por dificultad, comenzando por los más sencillos; ello permite a los jueces comprender con relativa facilidad la idea que subyace a los reactivos, es decir, que son un reflejo de los problemas a los que un estudiante puede enfrentarse y con ello establecer cuáles de ellos pueden ser respondidos por estudiantes con diversas habilidades. La tarea requiere que los jueces identifiquen un reactivo que marcador, mismo que servirá de referente para inferir que una persona con las competencias o habilidades mínimas pueda

contestar correctamente a los reactivos que anteceden al seleccionado. Empíricamente tener “las habilidades mínimas” es tener una probabilidad de 0.67 de contestar el reactivo correctamente (Karantonis y Sireci, 2006 y Lin, 2006).

El desarrollo de este procedimiento inicia con la capacitación de los jueces, posteriormente se hacen los juicios por rondas (normalmente se realizan tres rondas), lo cual se realiza para lograr una convergencia entre las opiniones de los jueces. En pequeños grupos los jueces discuten: a) qué conocimientos o habilidades están inmersos para contestar correctamente a los reactivos que se les presentan, y b) qué es lo que hace que cada reactivo dentro del cuadernillo sea más complicado que el que se ve previamente; inicialmente esta decisión se realiza de forma individual. Al finalizar la primera ronda se retroalimenta a los jueces considerando la mediana del conjunto de los puntos de corte. En la segunda ronda de discusión se analiza qué es lo que los estudiantes deben saber para alcanzar el nivel de desempeño establecido, con ello en mente se establece un segundo reactivo marcador en sustitución del primero. La tercera ronda comienza presentado los datos de impacto (es decir, el porcentaje de estudiantes en cada una de las categorías con base en los puntos de corte establecidos); este impacto se realiza en pleno con el fin de llegar a un consenso con todos los participantes. El método finaliza con la redacción de la descripción en prosa de los niveles de desempeño (es decir, lo que los evaluados son capaces de hacer) en función de los puntos de corte establecidos sobre el continuo de reactivos.

Para el EGEL este proceso se desarrolla mediante el juicio del CT, quienes en un primer momento observan los reactivos (previamente depurados mediante el uso de los indicadores psicométricos ya mencionados) en orden de dificultad, del más sencillo al más difícil, con el fin de que de forma individual cada juez defina el reactivo marcador que establece las diferencias entre los conocimientos requeridos para afirmar que un sustentante tiene un desempeño satisfactorio o sobresaliente. A partir de esta decisión se establece una ronda de impacto, en la cual se muestra al CT la distribución de sustentantes en cada uno de los niveles de desempeño (ANS, DSS y DSS) para cada una de las áreas de la prueba (que no se debe olvidar que en teoría representan las áreas de ejercicio profesional más representativas de la disciplina), lo que les permite a los jueces decidir hacer un ajuste o no sobre sus reactivos marcadores (que en conjunto definen los puntos de corte). Esta forma de trabajo permite la combinación de elementos técnicos y políticos para la definición de los puntos de corte, sin embargo, es un estándar que dependerá mucho del impacto

de la distribución de los estudiantes, más que de los contenidos y los mínimos establecidos por el juicio especializado de los miembros del CT.

Los criterios mínimos definidos para establecer si un egresado muestra un desempeño satisfactorio, sobresaliente, o bien, no alcanza el nivel mínimo permitirán la entrega de resultados. Como comentan Hambleton, Jaeger, Plake y Mills (2000), son categorías usuales en las pruebas criterioles para el establecimiento de estándares. Para tener un poco más de contexto en lo que respecta a esta cuestión vale hacer mención de las diferencias básicas en las pruebas normativas y criterioles. Las pruebas normativas permiten conocer la posición relativa de un sustentante en función de los demás, es decir, de todas las personas que presentan la prueba (Heredia, 2009). Es decir, el rendimiento de una persona estará en función del desempeño del conjunto de todos los participantes, tratando de medir las diferencias individuales con base en una curva de distribución normal, donde regularmente el estándar para “aprobar” es la media calculada por el conjunto de puntajes de todos los participantes. En contraparte, las pruebas criterioles pretenden “apreciar el logro de los objetivos por parte de cada alumno sin compararlo con el de sus compañeros” (Gómez Arbeo, 1990, en Heredia, 2009; pág. 43). En la evaluación criterial no es que las personas no se comparen, sino que su desempeño no se valora en función del resto de los evaluados, sino de un estándar en particular. Sin embargo, la definición de estándares en las evaluaciones de tipo criterial pudiera derivar en algunos problemas que los evaluadores debieran atender, tal como lo menciona Heredia (2009; pág. 46):

“Aquí el problema surge cuando a alguien se le ocurre que hay que mirar comparativamente los aprendizajes de los alumnos y en esa comparación se observan diferencias entre unos y otros, en realidad siempre las hay. Debido al ritmo individual de aprendizaje, unos logran antes que otros los objetivos, incluso unos logran el dominio del objetivo mejor que otros. Y entonces se considera que hay que establecer diferencias entre mejores y peores alumnos y reflejarlo mediante un tipo de evaluación objetiva adecuada, en este caso utilizando una escala de medida, pero de este modo nos situamos en la evaluación normativa”.

Dicho de otra forma, la evaluación criterial debe usarse con fines de retroalimentación, más que enfocarla en centrar diferencias entre estudiantes, ya que ello llevaría a hacer uso de la evaluación no con base a un criterio, sino con base a una norma. Por su parte, Ceneval, como se ha descrito, da un peso muy fuerte a los juicios de los miembros del CT en función del impacto de las decisiones de los reactivos marcadores; es decir, el CT observa en pleno una gráfica de distribución de las frecuencias de estudiantes en cada nivel de desempeño para valorar el impacto o la

proporción de estudiantes en ANS, DS y DSS, y con base en ello es posible ajustar el reactivo marcador que define los puntos de corte. Esto conlleva dos problemas, el primero de ellos relacionado con lo que comenta Heredia (2009) respecto a que las decisiones inicialmente consideradas como criterios se tornan normativas, ya que los puntos de corte dependerán en gran medida del impacto o de la perspectiva del CT respecto a la proporción de estudiantes en cada nivel de desempeño, por lo que los juicios sobre los estándares se ven fuertemente influidos por la distribución de los egresados en cada uno de los niveles de desempeño. El segundo problema se da al considerar un gran peso en la definición de un estándar la opinión de un conjunto de personas más que los contenidos (manifiestos en los reactivos depurados) que inicialmente se valoraron en el proceso del método Bookmark. Ambos problemas tienen consecuencias importantes, no solo debido al riesgo que se tiene de perder la perspectiva criterial que debe fundamentar al EGEL, sino que puede tener consecuencias en el ámbito pedagógico y político, ya que el alto impacto que el EGEL tiene dentro de la vida institucional y académica dentro de las IES está en función de una normalidad velada de criterio.

Adicionalmente se debe considerar que, como cualquier otro, el método Bookmark tiene algunas limitaciones. Lin (2006) realizó un balance del método mediante los criterios de Berk (1986), que son utilizados para valorar la pertinencia técnica y empírica de los distintos métodos para establecer niveles de desempeño; dentro de cada categoría se establece un juicio en una escala del 1 al 3, siendo 1 “no cumplido” y 3 como “cumplido completamente” para cada indicador. De acuerdo con Lin (2006), en el caso del método Bookmark cumple los aspectos de la siguiente forma:

**Tabla 4.3 Valoración del método Bookmark mediante los criterios de Berk.**  
*Pertinencia técnica*

<i>Indicador</i>	<i>Valoración</i>	<i>Justificación</i>
El método da información de clasificación apropiada	3	La metodología permite una clasificación relativamente sencilla de los estudiantes en los niveles de desempeño que se definan (por ejemplo, “capaz” o “no capaz”).
El método es sensible al desempeño de los evaluados	3	Es un elemento clave en el método, ya que el orden de los reactivos en los cuadernillos se determina en función de la dificultad de los reactivos, la cual se estima por las respuestas que los evaluados dan a cada pregunta.
El método debe ser sensible a la capacitación	3	Dado que son las propias respuestas de los evaluados las que determinan el orden de los

		reactivos, las respuestas están otorgadas por la enseñanza que los evaluados han recibido. Si los evaluados no han tenido experiencia académica con el contenido de la prueba, aumentará la probabilidad que se desempeñen de forma pobre en algunos reactivos.
El método debe ser estadísticamente robusto	3	El uso de los modelos TRI le da solidez estadística, basando sus estimaciones en el uso de la probabilidad.
El método debe identificar estándares verdaderos	3	El uso de grupos pequeños dentro de este método facilita las discusiones y el consenso. Si bien el método permite estimar un rango de error en el establecimiento del punto de corte (mediante el cálculo de medias de los juicios de los participantes), nunca estaremos seguros en dónde están los puntos de corte “verdaderos”.
El método debe permitir una decisión basada en evidencia	1	En general todos los métodos hacen uso de técnicas basados en poca evidencia, ya que para ello se requeriría un estudio longitudinal para identificar cómo se desempeñan en la realidad los evaluados clasificados en un nivel de bajo desempeño.

*Pertinencia práctica*

<i>Indicador</i>	<i>Valoración</i>	<i>Justificación</i>
El método debe ser de fácil implementación	3	Si bien requiere de mucha preparación para el ordenamiento de reactivos y llevar a cabo tres rondas de juicio, es sencillo de implementar.
El método debe ser sencillo de computar	3	Si bien en un inicio requiere del uso de programas estadísticos que permitan conocer los valores psicométricos de los reactivos utilizados, una vez que comienza el desarrollo del jueceo, la manipulación de los datos es sencilla.
El método debe ser de fácil interpretación para los lego	3	Si bien el método requiere de estimaciones matemáticamente complejas, los jueces en su carácter de lego en psicometría no requieren conocer a detalle cómo se desarrollan los cálculos para poder definir los puntos de corte.
El método debe ser creíble para los lego	3	La investigación sobre este método muestra que los jueces que lo utilizan suelen tener una buena comprensión del mismo.

Fuente: Lin (2006).

El método sin duda presenta muchos puntos a favor en lo que respecta a su solidez estadística y la manera amigable de trabajar con los jueces, sin embargo, también presenta algunas limitaciones. En primer lugar, dado que el método depende de las probabilidades de respuesta a un reactivo, la cual regularmente se estima en 0.67 para un evaluado con la capacidad teórica mínima, pero si esa probabilidad se modifica, entonces el orden de los reactivos puede cambiarse y alterar el resto del procedimiento. Otro problema es la “desordinalidad” de los reactivos, la cual refiere al desacuerdo entre los jueces sobre la forma en que los reactivos están ordenados (Skaggs y Tessema, 2001). Normalmente los jueces no están de acuerdo con la manera en que los reactivos están ordenados, ya que la dificultad empírica dada por los estudiantes no necesariamente coincide con sus perspectivas de los contenidos que en teoría debieran ser fáciles o difíciles; como consecuencia, el establecimiento de los puntos de corte entre jueces puede variar. La forma de intentar solucionar este problema ha sido discutir sobre lo que cada reactivo pretende medir y qué es lo que hace más fácil a los reactivos que le anteceden a la pregunta que se analiza.

Un problema adicional es que el único factor utilizado para tomar decisiones sobre los puntos de corte es la dificultad de los reactivos, sin considerar otros aspectos relevantes, como la importancia de los contenidos, por ejemplo. Por supuesto, también resulta relevante mencionar que las restricciones propias del modelo TRI limitarán la determinación de los puntos de corte<sup>20</sup>; por ejemplo, se asume que todos los reactivos tienen independencia local; si los supuestos de la TRI no se cumplen, el método entonces no es aplicable por completo a los reactivos analizados.

El uso del método Bookmark en el EGEL ha facilitado la inclusión de los miembros del CT en las decisiones sobre las formas de calificar o clasificar a los egresados. A partir del establecimiento de los puntos de corte se redactan los niveles de desempeño (ANS, DS y DSS), es decir, se describen en prosa las habilidades o conocimientos con los que cuenta un egresado cuyo desempeño se encuentra dentro de los límites de corte establecidos mediante el método; para ello el mayor insumo lo representan el conjunto de los reactivos que quedaron ubicados dentro de cada nivel de desempeño, ya que a partir de la descripción en prosa de lo que éstos pretenden medir es como se elaboran estos descriptores. Este proceso se realiza para cada una de las áreas de la prueba y

---

<sup>20</sup> Como cada modelo estadístico, el de la TRI tiene diversos supuestos que deben ser cumplidos para que el modelo sea válido para hacer juicios sobre los valores de los reactivos. Los supuestos son: a) La unidimensionalidad del rasgo latente; es decir, todos los ítems deben medir un solo rasgo en común. b) Independencia; las respuestas de cualquier evaluado a un reactivo son independientes de sus respuestas a otros reactivos (Cortada de Kohan, 2004).

una vez que se cuenta con los puntos de corte establecidos se procede a discutirlos en plenaria a fin de establecer los criterios para la asignación de *Testimonios de desempeño*.

Los testimonios de desempeño refieren a una forma de clasificar el dominio de los egresados sobre su disciplina considerando las áreas de cada prueba. Independientemente del EGEL, cada sustentante podrá obtener dos testimonios: Testimonio de desempeño satisfactorio (TDS) o Testimonio de desempeño sobresaliente (TDSS). Si un estudiante no cumple con ciertos requisitos en particular, puede incluso no recibir testimonio alguno. De acuerdo con Ceneval (2013b, pág. 17), “en todos los casos, dicho criterio refiere a una determinada cantidad de áreas en las que el estudiante debe alcanzar un resultado u otro”. Cada EGEL sigue una lógica de asignación de Testimonios en función de las decisiones del CT, por ejemplo, en el EGEL en Pedagogía se requiere alcanzar un nivel de desempeño satisfactorio o sobresaliente en al menos tres de las cuatro áreas que integran en el examen para obtener un TDS; por su parte, se requiere que el egresado alcance nivel satisfactorio en dos áreas y las restantes con desempeño satisfactorio para poder obtener un TDSS.

Si bien uno de los puntos más rescatables del EGEL es la de resaltar al aspecto cualitativo de su calificación, desde 1995 también se emite un puntaje conocido como el índice Ceneval (ICNE). Su escala abarca desde los 700 hasta los 1,300 puntos. Su cálculo se realiza de la siguiente manera:

$$ICNE = \frac{(PCNE_1)(Q)}{T_p} + Li_{ICNE}$$

- Donde:
  - PCNE<sub>1</sub>: Porcentaje de aciertos obtenido por el evaluado. Para el caso de puntuaciones por arriba del primer punto de corte se calcula la diferencia entre PCNE y PC1 o PC2, de acuerdo con el caso.
- Q: Posiciones existentes en cada intervalo de calificación
  - Valores: 1,000-700=300  
1,150-1,000=150  
1,130-1,150=150
- T<sub>p</sub>: Intervalo porcentual de puntos de corte
  - Valores: Tp1=PC1  
Tp2=PC2-PC1  
Tp3=100-PC2
- Li(ICNE): Límite inferior del intervalo de calificaciones del Índice Ceneval

- Valores: 700 (700-999)
  - 1,000 (1,000-1,149)
  - 1,150 (1,150-1,300)

Dado que en los EGEL se establecen dos puntos de corte para identificar tres niveles de desempeño (ANS, DS y DSS) en cada una de las áreas, los límites del ICNE se equiparan a los puntos de corte establecidos por el CT. De tal suerte que la obtención de un puntaje entre 700 y 999 siempre equivaldrá a un desempeño ANS, un puntaje entre 1,000 y 1,149 a un DS y un puntaje entre 1,150 y 1,300 a un DSS (Ceneval, 2013b), ello independientemente de la decisión del punto de corte establecido por el CT. Siendo así, los puntajes entre áreas y entre EGEL no son equiparables, ya que aunque la escala sea aparentemente la misma, ésta se ajusta a las decisiones de cada área de la prueba (recordando que los puntos de corte se establecen de forma independiente para cada área de un EGEL). Frente a ello, vale más considerar la interpretación cualitativa del desempeño de los estudiantes por cada área de forma independiente, ya que resulta metodológicamente incompatible realizar la suma de puntajes de todas las áreas para el establecimiento de un resultado en particular.

Todos estos elementos son los que se toman en cuenta en el proceso de emisión de calificación de los egresados. Si bien el método se basa en aspectos estadísticos robustos, tienen ciertas características que ponen en riesgo los elementos criteriosales que sustentan el quehacer del EGEL, ya que, como se mencionó, la consideración del conjunto de los resultados de los estudiantes en el establecimiento del impacto durante la determinación de los puntos de corte resulta un retorno a la consideración de la norma con el fin de, paradójicamente, establecer estándares para un juicio criterial. Adicionalmente, resulta riesgoso reducir todo el posible desempeño de un egresado en los criterios que actualmente tiene el EGEL, ya que, como comenta Stake (2006; pág. 117), “todo intento de resumir la calidad del rendimiento de un evaluando en un único criterio está prácticamente condenado a simplificar la evaluación en exceso”. Por supuesto, vale reconocer el logro de este proyecto al permitir el involucramiento de diversas opiniones de diversos agentes (como miembros del CT) en una decisión tan sensible como resultan ser las calificaciones de los egresados participantes en la evaluación.

## Emisión de resultados. ¿Cuáles son los elementos de decisión que soportan la forma de presentar resultados?

La emisión de resultados es un proceso fundamental dentro de cualquier evaluación, sin ella la evaluación carece de sentido. En el caso de una evaluación de egresados se debe considerar que detrás de cada resultado hay un profesionalista que recién terminó su formación y los resultados deben reflejar dicha sensibilidad. Como se comentó en los primeros capítulos de este trabajo, la evaluación puede emitir resultados con fines de uso político, de decisión o bien, con fines de retroalimentación. El EGEL, si bien inició como una pieza del rompecabezas pensado en retroalimentar el avance educativo del sistema de educación superior, centra su estrategia de emisión de resultados en dos sentidos: con un reporte individual y un reporte a las instituciones.

El reporte individual contiene los elementos de identificación del evaluado y la identificación de la IES a la que pertenece. El resultado aparece segmentado, en la parte izquierda se presenta si el egresado fue acreedor a un testimonio de desempeño (TDS o TDSS). En la parte de abajo aparecen los resultados por área en términos de nivel de desempeño alcanzado, así como su equivalencia en el Índice Ceneval. Asimismo, para poder facilitar la interpretación se muestran los criterios acordados por el CT para la obtención de un testimonio en particular. La siguiente figura muestra el anverso del informe individual de resultados del EGEL en Ingeniería Industrial.

**Figura 4.3. Ejemplo del anverso de reporte individual de resultados.**

REPORTE INDIVIDUAL DE RESULTADOS

Folio: 107122004

Nombre del sustentante: **RODRÍGUEZ DANTE EDUARDO EMILIANO**

Fecha de aplicación: 8 DE MARZO DE 2013

Institución de Educación Superior (IES): INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE MONTERREY-CAMPUS MONTERREY

Clave de identificación de la IES: 187965

<p>Díctamen general en el examen</p> <p><b>Satisfactorio</b></p>	Criterios para el otorgamiento del testimonio de desempeño en el examen	
	Testimonio de Desempeño Satisfactorio (TDS)	Al menos tres áreas con DS o DSS
	Testimonio de Desempeño Sobresaliente (TDSS)	De las cinco áreas, al menos una con DSS y las restantes con DS

Desempeño en cada área del examen					Criterios para determinar los niveles de desempeño por área	
Estudio del trabajo	Gestión de la cadena de suministro	Formulación y evaluación de proyectos	Sistemas productivos	Gestión industrial	Aún no satisfactorio (ANS)	700-999
DS	DS	ANS	DSS	DSS	Satisfactorio (DS)	1000-1149
1130	1122	977	1195	1211	Sobresaliente (DSS)	1150-1300

FIRMA DIGITAL:	<<<	38489369817A54E1A067C3D325D7CB64DF4B8884F0144860A8249E9DF36E2677 93946A833B982C6B421F7902A5B51F0B97188B946C584C1151F7F35A6K3D34A	>>>
----------------	-----	---	-----

Fuente: Ceneval (2013b). Los datos del sustentante e institución no son reales y son publicados directamente por la fuente consultada.

Por su parte, el reverso del reporte incluye los descriptores de desempeño redactados en la reunión para la definición de puntos de corte con el CT. Esta descripción, la cual está basada en los reactivos que se encuentran entre puntos de corte, pretende mostrar a los evaluados lo que éstos son capaces de hacer en cada área profesional dependiendo del nivel de desempeño obtenido en cada una de ellas. La siguiente figura muestra un ejemplo del reverso del informe individual de resultados del EGEL en Ingeniería Industrial.

Figura 4.4. Ejemplo del reverso de reporte individual de resultados.

		<b>Evaluar... Una buena medida para educar mejor</b>		<b>Examen General para el Egreso de la Licenciatura en Ingeniería Industrial EGEL-IINDU</b>	
NIVELES DE DESEMPEÑO POR ÁREA DEL EXAMEN					
Desempeño satisfactorio			Desempeño sobresaliente		
<b>Estudio del trabajo:</b> El sustentante es capaz de analizar, evaluar y proponer mejoras a los sistemas de producción vigentes, aplicando herramientas de la ingeniería de métodos y medición del trabajo, así como la aplicación de las técnicas propias de la ergonomía, higiene y seguridad industrial en el sistema hombre-máquina-medio ambiente.			<b>Estudio del trabajo:</b> Además de los conocimientos y habilidades del nivel de desempeño satisfactorio, el sustentante es capaz de utilizar las herramientas cualitativas y cuantitativas del estudio del trabajo para determinar el tiempo estándar de las operaciones de un sistema productivo, identificando sus consecuencias en los estándares de producción y en el factor humano, así como analizar las condiciones de sustentabilidad del proceso para proponer mejoras al mismo.		
<b>Gestión de la cadena de suministro:</b> El sustentante es capaz de identificar y resolver aspectos relativos a la cadena de suministro tales como pronosticar la demanda, identificar los sistemas de producción adecuados, calcular los requerimientos de capacidad de producción, programar la producción, selección del equipo adecuado y costo logístico.			<b>Gestión de la cadena de suministro:</b> El sustentante aplica sus conocimientos y habilidades del nivel de desempeño satisfactorio, para analizar, proponer y aplicar los métodos, modelos y/o sistemas adecuados para gestión de cadena de suministro tales como el cálculo de los parámetros de modelos de inventario, evaluar proveedores, planes maestros de producción, optimización de rutas y balanceos de flujo de productos.		
<b>Formulación y evaluación de proyectos:</b> El sustentante es capaz de utilizar las herramientas de la ingeniería económica con la finalidad de integrar alternativas de inversión.			<b>Formulación y evaluación de proyectos:</b> Además de los conocimientos y habilidades del nivel de desempeño satisfactorio, el sustentante es capaz de evaluar la viabilidad técnica y económica de un proyecto de inversión para la toma de decisiones.		
<b>Sistemas productivos:</b> El sustentante es capaz de determinar el área y espacios para distribución de planta y manejo de materiales que garantice la seguridad, higiene y el medio ambiente laboral, así como establecer los indicadores de productividad de la empresa a través de modelos parciales y totales para la optimización de los sistemas productivos.			<b>Sistemas productivos:</b> Además de los conocimientos y habilidades del nivel de desempeño satisfactorio, el sustentante es capaz de diseñar procesos de manufactura, identificando las características y requerimientos técnicos que conlleven a la selección de maquinaria y variables de operación a través del uso de diagramas de proceso; optimizar sistemas de producción mediante sistemas automatizados, manufactura flexible y sistemas de manejo de materiales; así como desarrollar los sistemas de planeación y programación de la producción y mantenimiento para garantizar la fiabilidad de los recursos.		
<b>Gestión industrial:</b> El sustentante es capaz de integrar conocimientos y herramientas de planeación estratégica, gestión ambiental, administración del capital humano, certificación de la calidad y control estadístico del proceso; para planear, implementar y controlar los recursos de la empresa.			<b>Gestión industrial:</b> Además de los conocimientos y habilidades del nivel de desempeño satisfactorio, el sustentante es capaz de innovar y mejorar los procesos de la empresa aplicando los conocimientos y herramientas de planeación estratégica, gestión ambiental, administración del capital humano, administración de la calidad total y control estadístico del proceso.		

Fuente: Ceneval (2013b)

El reporte de resultados permite así que el sustentante pueda identificar de forma cualitativa lo que es capaz de realizar, ello de acuerdo con el contenido de los reactivos utilizados para la definición de puntos de corte. De esta manera los descriptores de desempeño resultan ser un estándar de fácil interpretación mediante el cual se compara cada estudiante. Los resultados pretenden entonces legitimar el esfuerzo depositado en el desarrollo de la prueba con fines que el egresado identifique lo que es capaz de realizar; el sustentante entonces puede inferir que su ejecución en la prueba, ya sea en papel o por computadora, se compara con un estándar fijo y objetivo. Sin embargo se debe considerar que este estándar está íntimamente relacionado con los egresados que presentaron la prueba junto con él, dado que los puntos de corte (y por consecuencia sus resultados) y los descriptores de desempeño están implementados en gran medida en función de la observación de la distribución de los sustentantes en cada uno de los niveles de desempeño.

Entonces se establece un problema en la emisión de resultados, mismo que se agrava con el otorgamiento de un testimonio. Como se comentó, dependiendo de las condiciones implementadas por cada CT, se puede emitir un testimonio de desempeño. Cuando el egresado cumple con los criterios establecidos se le otorga un documento que testifica su desempeño (ya sea TDS o TDSS); la siguiente figura muestra un ejemplo de un TDSS:

**Figura 4.5. Ejemplo de TDSS.**



Fuente: Guía del sustentante del EGEL-Psicología. Disponible en: <http://www.ceneval.edu.mx>

La obtención de un testimonio no representa una forma de titulación, ya que eso es una atribución propia de las IES. Sin embargo, es sin duda una forma en que Ceneval testifica o da fe respecto al desempeño de una persona en las áreas más representativas de su ámbito profesional, de acuerdo siempre con la visión de los miembros del CT. Se testifica que una persona alcanza los estándares necesarios establecidos por el CT para considerarlo con un desempeño satisfactorio, sobresaliente, o bien, sin testimonio. Si bien los testimonios siempre han sido utilizados en el EGEL desde su inicio, a partir del año 2012 comenzó a entregarse el Premio Ceneval, el cual tiene como propósito “reconocer y estimular a los egresados de diversos programas de licenciatura que alcanzan un desempeño excepcional en estas pruebas de egreso, además de promover con ello la cultura del mérito y la excelencia académica en las instituciones de educación superior (IES) del país” (Ceneval, 2012b; pág. 1). Cada año se entrega una medalla a los sustentantes que alcanzaron un DSS en todas las áreas de cada EGEL, siempre y cuando hayan obtenido tal resultado en la primera ocasión que presentaban la prueba y la hayan contestado como máximo un año después de su egreso.

La emisión de los testimonios, y como consecuencia, ser acreedor al Premio Ceneval, si bien son un aliciente para los sustentantes, representan dos problemas fundamentales respecto a los propósitos de esta evaluación. El primero de ellos refiere a que es una medida centrada completamente en aspectos de acreditación. Paradójicamente se pretende promover la cultura del esfuerzo académico mediante un reconocimiento, incumpliendo así su función primaria definida por ANUIES (1993) como esencialmente constructiva y educativa, en lugar de la función meramente calificadora. La obtención de medallas son recompensas importantes para los egresados, sobre todo en una época de difícil inserción laboral, de tal suerte que la función de esta evaluación se ha adaptado para ser socialmente útil en un sentido pragmático. Esto se ha logrado a costa de una función formativa, modificando así el sentido que dio origen al EGEL, el cual al inicio se valoraba como la herramienta para aportar información con el fin de definir acciones para la mejora educativa a considerarlo ahora como un medio para estimular a los egresados y de promover con ello la cultura del mérito; sin embargo, ha favorecido su implementación como un medio de reconocimiento institucional y de estudiantes.

Si bien la evaluación formativa y sumativa no son aspectos antagónicos, no puede considerarse a una evaluación de este tipo solo como un elemento de titulación o administrativo. Una evaluación puede ser formativa y sumativa a la vez, ya que sus resultados podrán utilizarse para la mejora de un programa educativo (Stake, 2006), pero el EGEL no puede usar sus resultados con fines netamente de acreditación, que es un objetivo diferente al que le dio origen.

Adicionalmente en cada aplicación se presenta un Informe por IES de procedencia. En el anverso se presenta el puntaje obtenido por cada evaluado en las diversas áreas del EGEL, así como la información sobre si cada egresado fue acreedor a algún testimonio de desempeño; esto puede apreciarse en la siguiente imagen.

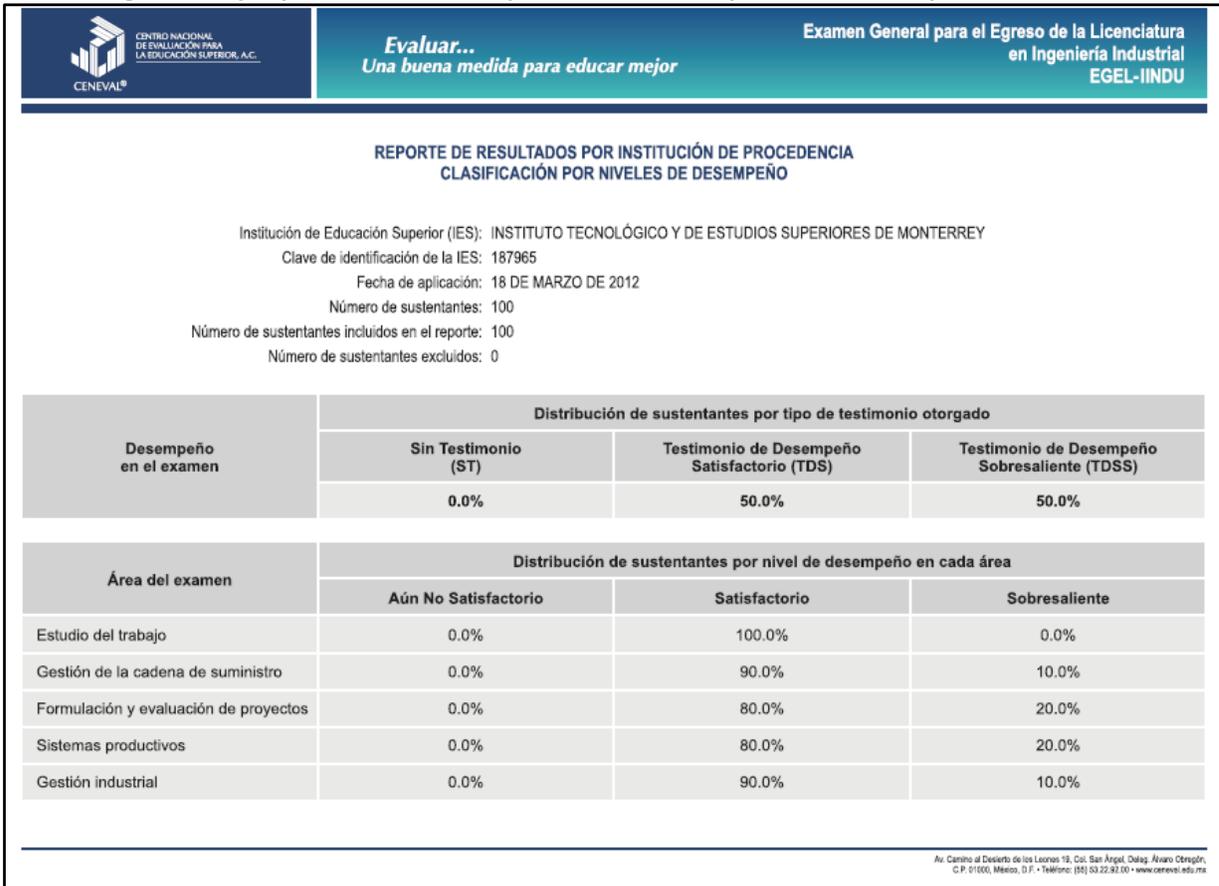
Figura 4.6. Ejemplo de anverso del reporte de resultados por institución de procedencia.

 CENTRO NACIONAL DE EVALUACIÓN PARA LA EDUCACIÓN SUPERIOR, A.C. CENEVAL®			Evaluar... Una buena medida para educar mejor					Examen General para el Egreso de la Licenciatura en Ingeniería Industrial EGEL-IINDU	
<b>REPORTE DE RESULTADOS POR INSTITUCIÓN DE PROCEDENCIA</b> <b>ORDENAMIENTO DESCENDENTE</b>									
Institución de Educación Superior (IES): INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE MONTERREY Clave de identificación de la IES: 187965 Fecha de aplicación: 18 DE MARZO DE 2012 Número de sustentantes: 3									
Núm.	Nombre del sustentante	Número de folio	Desempeño en cada área del examen					Testimonio de desempeño	
			Estudio del trabajo	Gestión de la cadena de suministro	Formulación y evaluación de proyectos	Sistemas productivos	Gestión industrial		
			NIVEL/PUNTAJE	NIVEL/PUNTAJE	NIVEL/PUNTAJE	NIVEL/PUNTAJE	NIVEL/PUNTAJE		
1	SANCHEZ ASENCIO OSCAR	258963251	DSS 1185	DS 1137	DS 1073	DSS 1205	DSS 1223	Sobresaliente	
2	BECERRIL LÓPEZ ALEJANDRA	125245212	DS 1130	DS 1122	ANS 977	DSS 1195	DSS 1211	Satisfactorio	
3	CASILLAS LIO ELÍAS	642512125	DSS 1170	ANS 897	ANS 885	ANS 904	DSS 1160	Sin Testimonio	
<b>Criterios para el otorgamiento del testimonio de desempeño en el examen</b>				<b>Criterios para determinar los niveles de desempeño por área</b>					
<b>Testimonio de Desempeño Satisfactorio (TDS)</b>		Al menos tres áreas con DS o DSS			<b>Aún no satisfactorio (ANS)</b>		700-999		
<b>Testimonio de Desempeño Sobresaliente (TDSS)</b>		De las cinco áreas, al menos una con DSS y las restantes con DS			<b>Satisfactorio (DS)</b>		1000-1149		
					<b>Sobresaliente (DSS)</b>		1150-1300		
Av. Camino al Desierto de los Leones 19, Col. San Ángel, Deleg. Álvaro Obregón, C.P. 01000, México, D.F. • Teléfono: (55) 53 22 92 00 • www.ceneval.edu.mx									

Fuente: Ceneval (2013b). Los datos de los sustentantes e institución no son reales y son publicados directamente por la fuente consultada.

Por otra parte, el reverso muestra el porcentaje de egresados en cada uno de los niveles de desempeño en las áreas que comprende el EGEL. La siguiente figura muestra un ejemplo de ello.

Figura 4.7. Ejemplo de reverso del reporte de resultados por institución de procedencia.



Fuente: Ceneval (2013)

Si bien el reporte por IES de dependencia permite tener una valoración global del desempeño de los egresados que presentaron dicha prueba en un momento en particular, reúne información centrada únicamente en el resultado final de una prueba, pero no debe perderse de vista que la evaluación es una actividad reflexiva que no solamente se ocupa de la valoración de objetivos al finalizar el programa (Cronbach, en Stufflebeam y Shinkfield, 1993). Incluso, desde un punto de vista institucional, la información centrada únicamente con fines sumativos presentada no permite observar posibles líneas de acción de mejora o, incluso, conocer las causas, contextos o situaciones en los que se encuentran los evaluados; ello debido a que la presentación de porcentajes aislados no representa un insumo que permita a los responsables institucionales y de programas educativos identificar cuáles son los elementos en los que se debe poner atención para la mejora de la formación de estudiantes. Si bien se presentan las áreas en las cuales los estudiantes mostraron un menor desempeño, no se explicitan elementos de la perspectiva

pedagógica que se encuentra en el trasfondo de los datos cuantitativos, o bien, no hay una descripción de las habilidades que es necesario desarrollar o los conocimientos teóricos que los estudiantes no dominan.

Para finalizar el análisis sobre la forma en que el EGEL comunica sus resultados, es necesario mencionar los informes anuales de los EGEL, los cuales son publicados cada año en el apartado “Estadísticas de los EGEL” en: <http://www.ceneval.edu.mx/ceneval-web/content.do?page=8308> En esta página web se encuentran reportes de los EGEL desde el año 2006. En estos informes incluyen: 1. Descripción del EGEL, 2. Datos descriptivos de la población que presentó la prueba, 3. Resultados del EGEL (en términos de cantidad de testimonios otorgados), 4. Relación de programas/campus que se incorporan al Padrón de Programas de Licenciatura de Alto Rendimiento Académico-EGEL, 5. Relación de sustentantes que obtuvieron el Premio Ceneval, 6. Variables de contexto y su relación con resultados del EGEL, 7. Opinión de los sustentantes acerca de la prueba. En particular, los puntos 2, 6 y 7 se obtienen mediante una breve encuesta socioeconómica que se aplica a los estudiantes al momento en el que se inscriben a la prueba.

Estos informes contienen información valiosa, no para una institución en particular (dado que los datos no muestran a qué institución pertenecen los sustentantes), pero sí para comprender la dinámica general del EGEL y buscar posibles relaciones de los resultados con variables como el promedio de licenciatura, el sexo o aspectos económicos. De cierta manera representa un valioso ejercicio de acercamiento sobre el contexto de los sustentantes y algunas causas que explicarían su rendimiento en el instrumento. Lamentablemente no se encontró evidencia publicada sobre el posible uso de las IES de esta información en revistas especializadas en educación o en informes de las propias IES usuarias o incluso en la página web del Ceneval, lo cual hubiera permitido valorar el posible impacto formativo del EGEL. Sin embargo, su posible nula consideración, así como la atención mostrada al Padrón de programas previamente mencionado, también refleja el uso de certificación que ha predominado de esta evaluación, la cual, como se ha comentado ya, puede acarrear consecuencias de índole político y pedagógico, corriendo el riesgo así de flexionar el objetivo primario de esta evaluación como pieza fundamental de la búsqueda de la mejora educativa de la educación superior en México.

## A manera de cierre

El presente capítulo mostró una descripción del EGEL como caso emblemático de la evaluación a gran escala en nuestro país, pero en particular como elemento de la evaluación del egreso a nivel mundial. El EGEL es un proyecto pensado desde los noventa con fines de ser una pieza de la estrategia de la búsqueda de la calidad de la educación superior en nuestro país como consecuencia de procesos sociales y académicos relacionados con la masificación de la educación superior en México y como un elemento para el establecimiento de una política de educación superior basada en rendición de cuentas, la cual lleva como consecuencia la búsqueda de certificación de conocimientos esenciales para la práctica profesional. Como se observó, comenzó bajo una evidente tendencia formativa aunque, paradójicamente, se le haya considerado únicamente como una *prueba*, lo cual impactaría en su metodología de diseño y en los resultados que se emiten a la sociedad, los cuales se centran fuertemente en la certificación de conocimientos y su uso como legitimación de las IES y del desempeño de sus egresados.

Siendo este el camino que se ha tomado con este proyecto, vale preguntarse si efectivamente es una evaluación o bien ha sido solo una herramienta certificadora que ha facilitado la reafirmación de la cultura de la rendición de cuentas, incluido en ello el desempeño de los egresados con fines de inserción laboral, así como la obtención de recursos a través de programas federales. Este último punto no es asunto sencillo, ya que la falta de coordinación del SES que se ha mencionado y la disminución de recursos públicos para la educación superior han favorecido una falta de control que conlleva a que las IES busquen elementos para sustentarse financieramente, dando así oportunidad a que la evaluación se utilice como medio para su obtención. El EGEL es un proyecto ambicioso que implica la participación de una gran cantidad de personas y una gran diversidad de perspectivas pedagógicas y metodológicas en la enseñanza de 40 distintas disciplinas, sin embargo, la diversidad no está del todo representada en su perspectiva epistemológica centrada en la atomización de conocimientos unificados o en la omisión de la diversidad curricular existente entre disciplinas.

El EGEL es un valioso esfuerzo que ha permitido tener avances técnicos importantes en la medición de aspectos de relativa transversalidad de contenidos entre programas de diferentes IES, sin embargo, es necesario pensar este esfuerzo considerando la complejidad de la realidad educativa en la cual se desenvuelven los egresados. Como comentan Jornet, Leyva y Sánchez (en Jornet y Leyva, 2009), es necesario considerar diferencias individuales y contextuales para emitir

un juicio de valor en un proceso de evaluación. Si bien es importante valorar los logros finales de los egresados, también es relevante valorar otros factores que hayan llevado al logro o no de los objetivos (Suchman, en Stufflebeam y Shinkfield, 1993). Ello representa un enorme reto para la evaluación del egreso, sin embargo pareciera un paso necesario frente a la complejidad de este fenómeno, así como los postulados que dieron inicio a este proyecto. Adicionalmente, no se debe dejar de lado el énfasis que institucionalmente se le da a la evaluación como elemento de mejora y así aumentar la calidad de la educación superior, pero que el uso de la evaluación del egreso como se ha hecho hasta ahora ha favorecido en gran medida a la mejora del control institucional y como elemento de obtención de recursos para el fortalecimiento de los programas educativos, reafirmando así las políticas educativas basadas en la rendición de cuentas. Vale también retomar la postura crítica de Díaz-Barriga (1994), quien afirma que si se piensa solo en acreditación, entonces se pervierte el sentido pedagógico del examen; en ese sentido, cuando las instituciones hacen uso de los exámenes como un instrumento que facilita la decisión respecto al ingreso o egreso y la discusión se centra únicamente en el cómo se realizan los instrumentos (es decir, la discusión versa solo sobre la metodología), entonces el único eje de análisis es la teoría de la ciencia y de la medida. De esta forma, si la discusión es técnica, la solución a los problemas pedagógicos se hace solo a ese nivel, haciendo así a esta actividad políticamente conservadora, sin fines de evolución educativa.

La amplitud de los elementos por considerar orilla al replanteamiento metodológico, que si bien el EGEL no hace una mención explícita de su modelo de evaluación, a lo largo de este trabajo se ha podido observar que ha sido fuertemente recargado en los postulados de Tyler. De acuerdo con Parlett y Hamilton (Stufflebeam y Shinkfield, 1993), la evaluación tyleriana presenta los siguientes problemas: la educación se caracteriza por elementos que no necesariamente pueden ser controlados; el evaluador debe poder adaptarse a situaciones cambiantes; los métodos tradicionales suelen dejar fuera datos relevantes para la evaluación, dado que solo se centran en los meramente cuantificables; los métodos estadísticos suelen ser menos sensibles a las particularidades locales; y dado que tradicionalmente se piensa en una *verdad objetiva*, se dejan a lado una diversidad de elementos de gran valía. Considerando así la enorme diversidad de modelos evaluativos, el próximo capítulo presenta una propuesta basada en el modelo comprensivo de Stake (2006), tomando como premisa fundamental que la evaluación no debe pensarse únicamente en su elemento técnico, sino que debe buscar una forma de comunicar elementos relevantes para la audiencia a la cual va dirigida (Stake, en Stufflebeam y Shinkfield,

1993). Este modelo favorecerá el acercamiento de la evaluación del egreso del nivel licenciatura hacia el uso formativo que permitirá que esta acción evaluativa se enfoque hacia una meta retroalimentadora que realmente favorezca la mejora de la formación de los próximos profesionistas mexicanos.

## **5. Propuesta de evaluación a gran escala del egreso de licenciatura bajo un enfoque comprensivo**

El capítulo previo permitió identificar los aspectos que conforman el EGEL como la herramienta de evaluación a gran escala del egreso de nivel profesional de México y uno de los ejercicios pioneros en nuestro país. La importancia de este proyecto recae en que esta experiencia de 20 años ha favorecido la generación de una cultura de evaluación dentro de las IES, por supuesto, ello en conjunto con otros esfuerzos, como los realizados por CIEES y COPAES.

A partir de este bagaje, así como en los elementos del enfoque comprensivo de evaluación (Stake, 2006), el presente capítulo pretende mostrar una propuesta metodológica respecto al desarrollo de una evaluación a gran escala del egreso de nivel licenciatura considerando una perspectiva alejada del enfoque tyleriano que ha definido a esta práctica desde hace 20 años en México.

La propuesta que a continuación se describe considera una organización similar a la presentada en el apartado previo, sin tomar en cuenta los antecedentes de la evaluación, la cual ha sido justificada en capítulos previos. La estructura de la propuesta es la siguiente: Epistemología de la evaluación a gran escala del egreso de la licenciatura, Metodología, Proceso de recopilación de información, Análisis de la información y Emisión de resultados.

### **Epistemología de la evaluación a gran escala del egreso de la licenciatura**

Como se ha comentado, uno de los aspectos más representativos proyectos de evaluación ha sido considerarlos como exámenes o pruebas, no solo en lo cotidiano, sino también por los mismos profesionales de la evaluación. Por supuesto, llevar a cabo una evaluación implicará indudablemente el desarrollo de algún instrumento o guía para el registro o medida de elementos que favorezcan la comprensión del fenómeno de interés y, por consecuencia, de comprender lo que se observa y emitir un juicio frente a un contexto en particular; ya que el conocimiento de un fenómeno (para su posterior valoración) no es posible sin un acercamiento al mismo. Es decir, dentro del proceso de evaluación, como lo es para la investigación y el conocimiento cotidiano, la relación epistemológica con el fenómeno es necesaria, en la cual el sujeto obtenga un conocimiento del fenómeno de interés. En el caso de la evaluación el método de acercamiento a ese conocimiento ha sido tradicionalmente de corte empírico, dado no solo la estrecha relación

que ha tenido el nacimiento de la evaluación con el desarrollo de métodos psicométricos de herencia del positivismo lógico, sino también debido a que el interés mismo de la práctica evaluativa refiere a hechos en concreto. Si bien el uso de este método es razonable dada la historia que ha caracterizado a la metodología de investigación psicológica y educativa, la concepción de una evaluación no debe verse reducida al instrumento que apoya su quehacer. Mantener esa línea de trabajo implica necesariamente subyugar un proceso epistémico del fenómeno complejo de la educación a la herramienta de trabajo que favorece una apreciación ingenua<sup>21</sup> de un elemento en particular de la educación mediante una relativa objetividad.

Si bien sigue una línea epistemológica positivista, la medición es relevante y necesaria en un proceso de evaluación, dado que es el proceso que permite asignar un valor a un elemento de interés, por lo que tradicionalmente ha sido el primer aspecto que se requiere para poder valorar el objeto de estudio. La medición se lleva a cabo a través de la observación, registros, o bien, solicitando a una persona que realice su propio recuento de lo que conoce u opina sobre un tema en particular, siempre con el objetivo de asignar un valor determinado al elemento de interés (normalmente expresado de forma numérica) que facilite nuestra comprensión del fenómeno. De acuerdo con Muñiz (1998), la medición es inherente a todas ciencias, aunque en las ciencias enfocadas a elementos humanos, como la psicología, antropología, sociología o educación, la medición conlleva ciertos errores difíciles de controlar, o dicho en otras palabras, implica la consideración de una aleatoriedad o aspectos que influyen en los valores emitidos y que evitan que éstos sean precisos. Este ha sido un aspecto discutido por psicómetras y solucionado con atenuantes como la calibración de reactivos, donde se ajustan los valores cuantitativos del instrumento mediante filtros de selección en función de los patrones de respuesta que los reactivos muestran dada una población en particular (como se explicó en el capítulo previo), proceso que se ve beneficiado de la consistencia de medida dada por la certeza el sustento teórico del instrumento y de las condiciones estandarizadas de aplicación. Si bien los procesos de calibración ayudan a la precisión de medida, es decir, a *filtrar* el instrumento, no logran concentrar

---

<sup>21</sup> El término *realismo ingenuo* es defendido por Guba y Lincoln (1985 en Ruiz, 2008) para referirse a la postura ontológica propia de la llamada *realidad objetiva* que supuestamente alcanza el método positivista. Se le denomina con el adjetivo *ingenuo* bajo una postura que indica que la objetividad no existe del todo, ya que la percepción de una realidad está supeditada a una postura, un contexto, así como a un conjunto de intersubjetividades de un grupo social, en este caso, la comunidad de especialistas encargada del desarrollo de la evaluación.

la complejidad que el fenómeno de lo humano implica, lo cual puede ser la raíz del llamado *error de medida*.

Ya se ha expuesto la manera en que la educación ha hecho uso de la psicometría para llevar a cabo sus evaluaciones, al considerar al aprendizaje como un rasgo psicológico. La psicometría es un elemento propio de la psicología científica (Muñiz, 1998), la cual pretende el estudio de la conducta humana y sus leyes. Al momento de considerar la postura bajo la cual se desarrolla una metodología hay que considerar lo comentado por Reichardt y Cook (1967, en Martinic, 1998) respecto a que un tipo de método se encuentra ligado irrevocablemente a un paradigma; de tal forma que si la evaluación que hace uso de elementos positivistas, tal como lo hace la psicometría, ello conlleva al uso de elementos cuantitativos para el cumplimiento de sus objetivos. Por supuesto, hay una gran confianza en los números debido a la relativa certeza que otorga favorece la valoración de lo observado. Sin embargo, con el desarrollo de los métodos de investigación, desde los años sesenta comenzó a implementarse una combinación de enfoques cuantitativos y cualitativos, y en los últimos 20 años se ha dado un fortalecimiento de este tipo de investigaciones, incluso, actualmente se ha optado por el desarrollo de triangulaciones para fortalecer la comprensión de un fenómeno educativo en particular (Pereira, 2011). En la actualidad el uso de diversos métodos para la recolección de información no se concentra únicamente en el uso de aspectos cualitativos y cuantitativos, o bien, a la triangulación de los mismos, al respecto Ruiz (2008; pág. 14) comenta lo siguiente:

“La transición paradigmática de la modernidad a la postmodernidad ha reanimado el viejo debate académico sobre los enfoques cuantitativo/cualitativo como métodos de investigación en ciencias sociales y educación. Desde el paradigma emergente se critican los supuestos metateóricos que fundamentan la investigación positivista, así como la instrumentación de la acción investigativa y el alcance de sus resultados en su intento de explicar la realidad social y educativa. En tal sentido, se cuestiona la validez epistemológica del conocimiento de lo social producido desde la óptica del positivismo lógico en la que, por una parte, se asume la relación sujeto-objeto como entidades independientes, y, por la otra, se cuestiona el estudio descontextualizado de la realidad, la concepción monológica de la misma, la visión analítica del objeto, la supuesta neutralidad y objetividad del investigador, la cuantificación de resultados, la excesiva formalización de los fenómenos y la pretensión de generar teorías explicativas de nivel universal.”

Si la investigación en educación comienza a trasladar su método hacia una postura epistemológica no centrada únicamente en análisis excesivamente analítico que consideran una relativa neutralidad frente al objeto, sino una apertura a un análisis a mayor profundidad, entonces la

evaluación de la educación y los procesos que están inmersos en ella (como el aprendizaje, la docencia, la didáctica, entre otras) serán más comprensivos para el contexto que las genera.

De esta forma, el interés de una evaluación a gran escala del egreso de nivel profesional, la cual tradicionalmente se ha visto a sí misma únicamente como un proceso de medición y que, paradójicamente, se ha mantenido ajeno a la complejidad humana que representa su propio objeto, debe centrarse en emitir una retroalimentación de los profesionistas considerando un contexto diverso, evitando mantenerse ajeno a una realidad intersubjetiva en donde las instituciones (con sus políticas y objetivos) y el contexto social se entrelazan en la vida de cada estudiante. En síntesis, no debe considerarse solo como un examen, a pesar de las ventajas en los indicadores de calidad y en financiamiento que ello pueda representar. El desarrollo de una evaluación centrada en el egreso de los nuevos profesionales no puede ser ajeno a la consideración de la complejidad humana en su quehacer. Hay que recordar que en las ciencias humanas (incluidas por supuesto, la educación y su evaluación) el conocimiento de un aspecto en particular siempre estará relacionado con la interpretación, la cual siempre será provisional y parcial de una realidad efímera (Contreras, s/f). Siendo así, los métodos en educación tendrán siempre una limitante, la cual debe asumirse como tal, evitando forzar los métodos e interpretaciones para generar una relativa objetividad de los procesos educativos. Por tanto, una evaluación de este tipo deberá reconocer que su visión y conclusiones siempre serán limitadas, no por cuestiones psicométricas en sí mismas, sino debido a que dada la complejidad y la ambigüedad propia del ser humano.

Si la evaluación del egreso considera entonces a la ambigüedad humana dentro de sí misma, no debe centrarse en la medición de rasgos de aprendizajes efímeros y descontextualizados y asumirlos como verdad única y absoluta, sino que debe considerar la que los resultados que se obtengan en una medición siempre estarán en relación a un contexto y a una formación didáctica determinada, lo cual implica abandonar la visión de enfocar la evaluación a un puntaje obtenido en un instrumento que no considera la diversidad y complejidad humana dentro de su construcción. Por lo tanto, la evaluación debe entonces enfocarse en los aspectos del contexto que dan cuenta de los resultados y no en los puntajes en sí mismos (por ejemplo, el plan y los programas de estudios), lo cual se logra modificando la conceptualización de la evaluación de la medición hacia la retroalimentación.

Partiendo de este análisis, la perspectiva que fundamenta la presente propuesta se basa en un enfoque comprensivo. Dicho enfoque significa una interpretación personal, vivir la experiencia de *estar ahí* (Stake, 2006). Considerar a la experiencia subjetiva como valiosa implica una fuerte ruptura con la tradición del positivismo lógico heredada de los métodos psicométricos que han mantenido la hegemonía de la medición sobre la retroalimentación, además que guarda estrecha relación con las tendencias de la complejidad en educación (también llamada *pedagogía de la complejidad*, Brower, 2009). En ese sentido, la investigación y la práctica educativa de los últimos tiempos ha virado de los métodos tradicionales hacia la consideración de la diversidad como fundamentación para la comprensión de una realidad compleja; al respecto, Colom (2005; pág. 3) lo expresa de la siguiente manera:

“Hoy de la educación se precisa, fundamentalmente, capacidad para enfrentarse a realidades complejas, relacionadas incluso por redes de sistemas, por un reticulado inextricable, complejo y dinámico que forma la sociedad actual que, no se olvide, es la sociedad de la globalización, de la mundialización y de las nuevas tecnologías. Estamos convencidos de que se necesitan nuevas formas de pensar la educación en un mundo asimismo nuevo, que nada tiene que ver con el de la modernidad. Nos enfrentamos a los sistemas complejos con herramientas intelectuales e heurísticas de otros tiempos, bajo una mentalidad evolutivo-lineal que contempla un mundo homogéneo, estable, en el que las mismas causas producen, más o menos, los mismos efectos.”

Por tanto, la evaluación de uno de los elementos del proceso educativo, en este caso, el egreso del nivel licenciatura, deberá tener en consideración la complejidad del fenómeno mismo, de otra forma representa mantener una perspectiva lineal en su desarrollo. Si bien ello implica la aceptación de una posible incertidumbre en los resultados, la no negación de la misma favorecerá para prescindir de la simplicidad de los métodos tradicionales, orillando a la consideración de la complejidad dentro de su desarrollo y ejecución. Retomando a Burgess (2004, en Oliver y Hyun, 2008) y a Colom (2005), la complejidad no deben entenderse como un desorden (en el sentido coloquial de la palabra), sino como una forma de aceptación que los procesos educativos son dinámicos y contextualizados, así como la inclusión de diversos actores en su ejecución, además de la consideración de numerosas perspectivas para el conocimiento y entendimiento de un fenómeno en particular. En la epistemología de la complejidad (Morin, 2004), no solo se debe considerar la existencia de diversos elementos constituyentes de un fenómeno en particular, sino que están conectados y dependen el uno del otro; de esta forma, es inviable pensar en los elementos en aislado.

La búsqueda constante de certidumbre que ha caracterizado a la evaluación del egreso a gran escala facultada únicamente por la medición no ha cambiado conforme ha venido mutando la sociedad y el trabajo teórico en la educación, por ende, representa un esquema que no corresponde del todo con la complejidad inmersa en los procesos educativos. Un modelo basado en la epistemología de la complejidad, como el que se propone ahora, rompe con el pensamiento lineal y lo reemplaza con enfoques más completos. Lo fijo en el ambiente no existe, la estabilidad es la *estabilidad de los muertos* (Morrison, 2003); es decir, para mantener un desarrollo, es necesario el cambio, y es el cambio el enfoque que debe permanecer en la evaluación de los futuros profesionistas. Este primer elemento es denominado *flexibilidad*. Adicional a él, es necesario que la evaluación tome en cuenta a la *diversidad*; la cual implica la consideración de diversos puntos de vista y una gran variedad de problemas y perspectivas en la enseñanza, y es que el modelo actual de evaluación a gran escala del egreso asigna bajo un criterio fijo los elementos que deberán evaluarse, excluyendo así ciertos aspectos que pudieran ser valiosos para valorar el pensamiento crítico de los estudiantes. Un tercer elemento es la *adaptabilidad*, el cual, en gran relación con la flexibilidad, la evaluación debe considerar acomodarse a diversas formas de aprendizaje; una evaluación debe considerar que las personas no son estáticas y su esencia es el cambio. Por otra parte, la evaluación debe enfrentar la *complejidad*, elemento característico de la sociedad actual, la cual implica una interconectividad entre disciplinas, culturas y una gran cantidad de información disciplinar que se produce día con día. Por último, se debe considerar la *conectividad*; bajo la teoría de la complejidad, todo está relacionado, la escuela no es un agente aislado y cada persona se relaciona con diferentes tipos de conocimiento. Es necesario entonces asumir el principio de incertidumbre dentro del conocimiento; si bien ello no obliga a cualquier persona a dominar enteramente todos los elementos y disciplinas relacionadas con su objeto de estudio, implica ser consciente que en el desarrollo de un “sistema de ideas” (Morin, 20014, pág. 13), la incertidumbre en su conjunción estará presente.

De acuerdo con Morin (1999), hay que sustituir la visión de un mundo lineal con un orden impecable a uno donde hay una dialógica entre el orden, el desorden y la organización. Ello no implica que el trabajo que se realice bajo esta perspectiva no guarde una estructura en particular, e incluso que no se consideren valiosos los procesos de medición, sino que nos invita a participar de una perspectiva de trabajo que asuma la incertidumbre *cerebro-mental*, que implica la reconstrucción de todo conocimiento, la incertidumbre *lógica*, ya que “ni la contradicción es señal de falsedad ni la no contradicción es señal de verdad” (Pascal, en Morin, 1999, pág. 42), la

incertidumbre *racional*, ya que incluso si lo racional no se juzga a sí mismo se vence a sí mismo, y la incertidumbre *psicológica*, ya que es completamente imposible identificar todo lo que sucede en la mente de alguien, incluso en la propia. Considerando esta visión, la perspectiva evaluativa a gran escala del egreso de la licenciatura no habría centrarse entonces en la medición misma como eje de su propuesta, sino en su conocimiento del fenómeno por analizar: el egreso de los nuevos profesionistas del país, con la diversidad que implica la multiplicidad de formaciones en nuestro SES, así como la presentada por la medición misma del aprendizaje.

El enfoque comprensivo de evaluación propuesto por Stake (2006) no considera de forma manifiesta una relación con la postura epistemológica de la complejidad, dado que el autor no hace explícita una argumentación ontológica ni epistemológica de su perspectiva; sin embargo, la relación del enfoque comprensivo con la postura epistemológica de la complejidad es clara. De acuerdo con Stake (2006), la evaluación debe considerar que el evaluando es único, por lo que sus resultados ni conclusiones pueden ser generalizados. También se entiende que la evaluación implica enfrentarse a cuestiones y prácticas humanas, por lo que es necesario entenderlas como acciones contextualizadas y mediante ello reconocer que se trabaja con experiencias vividas y no solo con aspectos conceptuales propios de una formación disciplinar. En ese sentido, la propuesta aceptará el dinamismo y la expansión constante de la formación profesional (Brower, 2009).

La evaluación comprensiva, de forma consistente con la epistemología de la complejidad, implica la consideración de la diversidad de perspectivas personales, dado que es parte del estudio de la actividad humana. Por ende, para que una evaluación pueda ser considerada como comprensiva es necesario que tome en cuenta los intereses y preocupaciones de todos aquellos que estén involucrados, debido a la diversidad de perspectivas para conformar una realidad implicadas en los procesos educativos y de evaluación. Es necesario que la evaluación a gran escala del egreso de la licenciatura incorpore a sí misma la diversidad de perspectivas, ya que, como comenta Santos Guerra (1988), la evaluación es un instrumento de poder tal que permite la construcción de una realidad en particular sobre una persona; siendo así, la construcción de la realidad sobre una persona no debe ser decisión del evaluador, sino de la diversidad de perspectivas y contextos que le rodean. Incluso, se vuelve imprescindible dar significado a una evaluación considerando para ello las claves de interpretación del contexto, de otra forma sería vaciar a la evaluación de la realidad misma (Oates, 1975, en Santos Guerra, 1988).

Como se mencionó en el Capítulo 2, una evaluación a gran escala del egreso de la licenciatura que se desarrolle bajo una metodología comprensiva no se debe centrar en juzgar, es decir, en emitir una conclusión de valor en función de un puntaje, sino en proporcionar información que facilite la toma de decisiones para mejorar al evaluando, en este caso, la formación de profesionistas en función de su egreso. Para ello se requiere forzosamente evitar a la evaluación como un elemento netamente técnico, cuyo interés se enfoque principalmente en lo metodológico, sino que se debe valorar en mayor medida un proceso dialógico con los *stakeholders* que favorezca el establecimiento de lo que se pretende evaluar. De esta manera la diversidad de *stakeholders* favorecerá la consideración de una multiplicidad de perspectivas sobre, haciendo uso de los conceptos de Stake, la filosofía del existir del evaluando, es decir, conjunta panoramas respecto a lo que el evaluando es en un momento dado en particular.

La metodología por utilizar privilegia también la comprensión de las características del evaluando, implicará entonces no dejar a un lado cuestiones propiamente humanas, más allá de únicamente los resultados educativos, tal y como las decisiones éticas que los egresados enfrentarán en su vida profesional. Por supuesto, estos elementos deberán valorarse de forma contextualizada.

La evaluación a gran escala del egreso de la licenciatura bajo este enfoque, además de considerar la diversidad de contextos, de prácticas educativas y de perspectivas sobre el egreso mismo, su eje debe ser entonces la retroalimentación a las instituciones y a los egresados sobre lo evaluado, todo ello mediante la recolección de diversas fuentes de información, lo cual no implica dejar a un lado la relevancia de la medición para poder acercarse al evaluando, pero sí es necesario no considerarla como determinante de su desarrollo. En ese sentido, la estrategia metodológica de la medición de rasgos psicológicos (por ejemplo, el aprendizaje o la opinión) no puede dejarse de lado dentro de una evaluación, dada la necesidad de observación de elementos empíricos, sin embargo, no es un aspecto que vaya en contra de la postura comprensiva, solo que debe ser considerada bajo una lógica distinta. Por supuesto la medición, mediante pruebas y cuestionarios, es lo que permitirá el acercamiento al fenómeno del egreso de la licenciatura, sin embargo, no deberá ser el eje de la propuesta y la metodología deberá aceptar que el fenómeno puede tener contradicciones (Morin y Pakman, 1994), además que el proceso de “flujicidad<sup>22</sup>” (Brower, 2009, pág. 80), no debe detenerse; es decir, no deben buscarse certezas definitivas con una aplicación

---

<sup>22</sup> Refiere a la introducción constante de información a los procesos de aprendizaje, es decir, a no considerar un momento educativo observado como definitorio. De esta forma la pedagogía de la complejidad invita a mejorar constantemente los métodos utilizados y a evitar las conclusiones y certezas definitivas.

de un instrumento, sino que se deberá alentar un constante acercamiento al fenómeno educativo en cuestión. Por supuesto, una metodología comprensiva implica también la necesidad de considerar la diversidad de agentes inmersos en el egreso de los estudiantes y tener siempre en cuenta las limitantes de juicio que la evaluación puede presentar, al no establecerse como verdad universal. El siguiente apartado muestra una propuesta de desarrollo con base en esta postura.

## Metodología

El desarrollo metodológico de cualquier práctica académica considera forzosamente su marco epistemológico como núcleo de acción. En el caso del presente trabajo, se considera como eje rector la perspectiva de la evaluación comprensiva. Siendo así, a continuación se describen los elementos para el desarrollo de una evaluación a gran escala del egreso de la licenciatura bajo este enfoque. La propuesta se divide en apartados que representan las fases o pasos metodológicos que en dado caso pudieran guiar este ejercicio. Cabe considerar que si bien la tradición de la evaluación a gran escala del egreso de la licenciatura se ha desarrollado bajo una metodología unificada y considerando únicamente elementos de la formación profesional como aquellos relacionados con sus contenidos, en esta propuesta los elementos de cada formación profesional, así como de contexto, deben considerarse de forma transversal. En cada una de los apartados se detallan elementos al respecto.

### *Establecimiento del objetivo de evaluación*

Uno de los aspectos más complicados en el desarrollo de cualquier proyecto de evaluación es uno de los más cruciales para su ejecución: la definición del objetivo. Aquello que persigue una evaluación no debe ser dado por hecho, y es que a partir de este elemento es que se define la visión de la evaluación, la construcción de los instrumentos, hacia dónde va dirigida, sus usos y formas de comunicación de sus resultados. Dentro de la cultura educativa de actualidad, basada en la rendición de cuentas, pensar en una evaluación implica casi de forma inmediata considerar la valoración los logros obtenidos una vez dada una inversión de recursos y esfuerzos. La reflexión del por qué evaluar se da poco, ya que en la nueva cultura educativa, la evaluación es necesaria *per se* frente a todo proceso educativo. Por supuesto, ésta es una perspectiva que puede aportar elementos positivos a los procesos educativos, siempre y cuando se considere no solo como una fuente de información necesaria para tomar decisiones de índole política y administrativa, sino también como elemento pedagógico para mejorar la formación de los estudiantes; sin embargo, el

esfuerzo de definir el objetivo de una evaluación de este tipo nos permitirá tener presente el camino metodológico por seguir.

Si bien la experiencia mexicana de la evaluación del egreso de la licenciatura a gran escala tiene por objeto prioritario la rendición de cuentas, como se observó en el Capítulo 2, las experiencias internacionales permiten conocer que existe una diversidad de objetivos (reflejados en el uso que se hace de los resultados) de este tipo de evaluaciones, desde la retroalimentación del currículo, análisis de tendencias del comportamiento de la matrícula y la evaluación de los programas de estudio. Ninguna evaluación debiera llevarse a cabo asumiendo un objetivo *per se*, sino que la evaluación y sus instrumentos deberán desarrollarse planteando un objetivo claro en congruencia con una postura epistemológica y un modelo de evaluación en particular. Por supuesto, es un elemento que no puede elaborarse en aislado, ya que debe conjuntar la definición de los aspectos por evaluar; de otra forma, el objetivo de evaluación estaría redactado sin conocimiento de lo que se pretende evaluar. Con base en ello, a continuación se presenta una propuesta de definición de los elementos por evaluar que permite la posterior definición del objetivo de la presente propuesta.

#### *Definición de elementos por evaluar*

Como se ha comentado, la propuesta actual no considera como eje rector el desarrollo de un instrumento de medida; sin embargo, tal como en los procesos de investigación educativa, es necesario considerar formas para la recolección de información que faciliten el conocimiento del egreso del nivel licenciatura mediante la medición de los rasgos psicológicos que se pretenden. En ese sentido, lo más relevante para la emisión de una retroalimentación no son los instrumentos en sí (aunque son relevantes al momento de considerar el elemento operativo que una levantamiento de información a gran escala implica, lo cual se aborda más adelante), sino los aspectos que deben delimitar la evaluación, es decir, preguntarnos *¿qué se debe evaluar?*

Como se ha abordado, tradicionalmente una evaluación a gran escala se centra en aspectos de aprendizaje o algún otro proceso cognitivo relacionado con la escolaridad o en función de ésta. Como en todo proceso de evaluación, es necesario definir los límites que definen al evaluando. El evaluando en este caso es el profesional recién egresado de nivel licenciatura en México. Siendo así, cabría preguntarse, *¿Qué se espera de un egresado de este nivel de estudios?* Hay que recalcar que la pregunta resulta ser el eje de desarrollo de la propuesta metodológica en conjunto con la

perspectiva epistemológica previamente definida. Considerando la pregunta mostrada, vale dar un giro a su perspectiva, con fines de alejarse de la postura tyleriana que se enfoca en los logros obtenidos al finalizar una formación educativa. Preguntar sobre *lo que se espera de un egresado* nos llevaría a indagar sobre el estándar que se establece como ideal al finalizar su educación; es decir, qué objetivos o qué características debe cumplir al final de una formación en una profesión en particular. Esta pregunta no solo presenta una dificultad dada su relación con el enfoque tyleriano que llevaría a hacer una propuesta de la misma línea epistemológica tradicional, sino que también orillaría a asumir que *lo que se espera de un egresado* es una situación universal e inmutable.

En el marco una perspectiva comprensiva, la pregunta no debe ir dirigida a los resultados, y no es que se excluyan de lo que se pretende evaluar, sino que se consideran junto con otros elementos. Bajo este análisis, la pregunta que nos permitirá la definición del evaluando es: *¿Cuáles son las características con las que cuenta un egresado de nivel licenciatura que le permitirán desempeñar su profesión de forma adecuada?* Una lectura rápida de la pregunta propuesta indica una aparente visión tyleriana, al centrarse en las *características con las que cuenta un egresado*; lo cual puede orillar también a la determinación de un estándar único. Para ahondar más en la manera en que la pregunta está planteada es necesario tomar en cuenta las siguientes aportaciones que la literatura nos ofrece sobre los egresados de nivel superior.

Actualmente la visión respecto a lo que representa ser un egresado de educación superior en México y en el mundo ha tenido diversos giros respecto a lo que tradicionalmente representaba. La idea tradicional del profesional competente que se basaba en la posesión de conocimientos y habilidades para poder desarrollarse con éxito es una perspectiva que ha quedado atrás (González y González, 2008). Siguiendo a Delors (1996, en González y González, 2008), en la vida profesional no basta con conocer y saber hacer, es necesario *ser un profesional*. Siendo así, ¿qué es ser un profesional? La palabra “profesión” proviene del latín *professio*, que significa acción y efecto de profesar (Fernández, 2001). El uso común se relaciona con el empleo u oficio que cada persona tiene y ejerce de forma pública. Actualmente las profesiones están íntimamente relacionadas con la formación en instituciones de nivel universitario, aunque la formación profesional universitaria tiene sus antecedentes en las universidades medievales, que se conjunta con la terminación del sistema feudal, así como el incremento de intercambios culturales y comerciales. La necesidad de una formación profesional se vio así relacionada con las exigencias demandadas para el desarrollo

de una sociedad ligada a los gremios de la Edad Media (Barrón, 1996, en Fernández, 2001). Desde un punto de vista formal, se requería contar con un *licentia ubique docendi* o un reconocimiento universal de los grados obtenidos y así poder ejercerlos en cualquier parte del mundo.

Siendo así, ser un profesional implica que el Estado participe en la regulación de las profesiones, normalmente en función de sus intereses. En ese sentido, “una profesión es una comunidad autorregulada que tiene un poder exclusivo, normalmente respaldado por el Estado, para adiestrar a nuevos miembros o admitirlos en sus prácticas, reservándose el derecho de juzgar la actuación profesional de sus propios miembros” (Fernández, 2001, pág. 8). Esto tiene estrecha relación con la función reguladora de la evaluación del egresado que previamente se analizó; como consecuencia, bajo esta dinámica no es posible separar la noción de profesional con el contexto y políticas del Estado respecto a lo que se espera de un profesional. De forma complementaria a esta visión, según Hortal (1996, en López, 2013) para que un trabajo sea considerado profesión debe tomar en cuenta: que sea el medio de vida de una persona, que exista un cuerpo específico de conocimientos de dicha actividad, que los conocimientos se transmitan de forma institucionalizada, que exista un control de los profesionales sobre el ejercicio de la actividad. Adicionalmente, cada profesión tiene una función social, como tal, cada profesión pretende dar un bien a la sociedad; a ello se le llama “bienes internos” de una profesión (Martínez, 2006, en López, 2013).

Cortina (2000, en López, 2013) indica que una profesión tiene como finalidad proporcionar a la sociedad un bien específico indispensable para su supervivencia como sociedad, por lo que se requiere que los profesionales se identifiquen con la sociedad. Siendo así, toda profesión tiene un fundamento ético, y, por tanto, toda acción profesional tiene que cumplir un deber ético. El elemento ético del profesionista no debe entenderse como especulaciones filosóficas que es necesario hacer a un lado bajo la mirada del Estado centrado en la rendición de cuentas y en objetivos. Si bien la visión científicista y racionalista dejó a la ética fuera del campo de lo científico (López, 2013), una mirada comprensiva del profesionista recién egresado no puede evitar considerar este componente como parte de la praxis humana, en particular si el ejercicio de definición del evaluando implica tomar en cuenta que toda profesión se ejerce desde una visión ética y toda visión ética conlleva una definición de profesión y de lo que es un *buen profesional*. Ahora bien, si la visión ética del ejercicio profesional no resulta solo en una especulación filosófica, sino un eje del comportamiento y desempeño profesional, es necesario considerar los principios

que lo sustentan. En primer lugar, el principio de *beneficencia* implica desarrollar la profesión de manera competente y eficaz, pensando siempre en el beneficio de los usuarios y de la sociedad. Esto implica la consideración de los efectos de sus decisiones sobre las personas o la sociedad. El segundo principio, el de *autonomía* busca aclarar que el beneficiario no es un agente pasivo, sino que debe participar en las decisiones que implica la prestación de un servicio profesional. De esta forma, un profesional ético debe considerar siempre a los usuarios de sus servicios como sujetos de derechos, poseedores de una dignidad inalienable y capaz de participar en la toma de decisiones sobre lo que les va a afectar. El tercer principio, el de *justicia* considera que el ejercicio profesional está enmarcado en un sistema social, por ende, cada sujeto involucrado debe cumplir con su deber. Siendo así, el principio de justicia se cumple cuando los profesionistas se preguntan por su contribución al bienestar general de la sociedad a partir de una adecuada organización institucional y normativa.

Los tres principios de ética en un profesional invitan a una reflexión respecto al espectro de posibilidades que se pueden considerar en la evaluación del egreso de la licenciatura, ya que conforman un fuerte sustento de lo que implica el profesionalismo. Por supuesto, si bien la consideración de los principios éticos de un profesional es el inicio de un modelo más amplio para la evaluación del egreso, no se debe olvidar que, como comentan Fernández (2001) y González y González (2008), la cultura profesional actual considera la necesidad de tomar en cuenta los avances tecnológicos, las políticas internacionales, comunicaciones, necesidades del sector productivo y requerimientos de la sociedad en general; sin olvidar los conocimientos y habilidades adquiridos, motivaciones y desempeñarse de forma eficiente en varios contextos.

Bajo esa perspectiva, además de los elementos éticos que pueden orientar la práctica profesional, sin duda es necesario tomar en cuenta las competencias que están inmersas en ello. Al respecto, la UNESCO (1998 en González, 2006), establece como una misión esencial de la educación superior la formación de profesionales competentes y comprometidos con el desarrollo social. Por supuesto, esta visión conjunta el elemento ético, así como el quehacer manifestado en la competencia. Ambos aspectos son parte del desempeño profesional de un egresado de educación superior, ya que, como comenta Bolívar (2005), la formación de un profesional competente no se puede desarrollar sin considerar la formación ética; y es que la profesionalidad comprende, además de competencias teóricas y/o prácticas, integridad personal y conducta profesional ética.

El término de *competencias* es altamente utilizado en el ámbito educativo en las últimas décadas y se ha buscado su implementación en los planes de estudio de formación superior, sin embargo, en México se ha avanzado muy poco al respecto, ya que no hay una claridad de lo que implica este modelo (Moreno, 2009). La búsqueda de la implementación de las competencias en la formación profesional tiene como antecedente su uso en el ámbito empresarial, como una necesidad de realizar predicciones acerca de la ejecución de las personas en su desempeño laboral (McClelland, 1973, en González, 2002). Las competencias laborales son un “Conjunto de características de una persona que están relacionadas directamente con la buena ejecución en una determinada tarea o puesto de trabajo” (Boyatzis, 1981, en González, 2002; pág. 1), o bien, una “característica subyacente de un individuo que está causalmente relacionada con un rendimiento efectivo o superior e una situación o trabajo definido en términos de criterios” (Spencer y Spencer, 1993, en González, 2002; pág. 2). Por su parte, Ansorena Cao (1996, en González, 2002), comenta que la competencia es “una habilidad o atributo personal de la conducta de un sujeto que puede definirse como característica de su comportamiento y bajo la cual el comportamiento orientado a la tarea puede clasificarse de forma lógica y fiable”, mientras que Rodríguez y Feliú (1996, en González, 2002; pág. 2) definen a la competencia como el “conjunto de conocimiento, habilidades, disposiciones y conductas que posee una persona que le permiten la realización exitosa de una actividad”. Por su parte, Vargas (2001, en González, 2002; pág. 2) las define como características permanentes de las personas, se ponen de manifiesto cuando se ejecuta una tarea o un trabajo, están relacionadas con la ejecución exitosa de una actividad, tienen una relación causal con el rendimiento laboral, pueden ser generalizadas a más de una actividad, combinan lo cognoscitivo, lo afectivo y lo conductual. Cada definición considera elementos propios, sin embargo, todas ellas recaen en un conjunto de habilidades y conocimientos que se ejecutan con el fin de realizar con éxito una tarea en particular. Cabe resaltar que las definiciones mostradas acentúan el hecho que las competencias son elementos subyacentes de las personas que deben ser manifestadas en una tarea, además que son permanentes. No quiere decir que sean estables (es decir, pueden desarrollarse), pero una vez que son parte del *repertorio* del sujeto se mantienen con él.

Complementando esta visión, Moreno (2009) expone una diversidad de perspectivas sobre el término (considerando a Norris, 1991; OCDE, 2005; Ebeling, 1992; Dale e Iles, 1992; Hartle, 1995; Elkin, 1990; Cockerill, 1989; Boak, 1991; Burgoyne, 1988; Woodruffe, 1991; Mansfield, 2004; Weinert, 1999; Perrenoud, 2004; Boon y Van der Klink, 2002), aunque todas ellas refieren a la competencia como un elemento de ejecución, así como la consideración de habilidades cognitivas

y no cognitivas y, en algunos casos, centrados en la solución de elementos profesionales o enfocados en demandas organizacionales. Como parte del desarrollo de investigaciones y evaluaciones enfocadas en generar conocimiento respecto a las competencias y a su evaluación, destaca el proyecto *Definition and Selection of Competences* (DeSeCo) promovido por la OCDE (Rychen y Salganik, 2003; Hernández y Rodríguez, 2008). DeSeCo fue un proyecto de evaluación de competencias que surgió como complemento a otras evaluaciones, como PISA. Implicó un trabajo internacional de definición teórica de competencias, en particular aquellas llamadas *competencias clave*, que de acuerdo con este proyecto son aquellas que pueden aplicar en diversos contextos y apoyan el desarrollo de una vida exitosa. Las competencias clave enmarcadas en este proyecto refieren a interactuar en grupos heterogéneos, actuar de forma autónoma y usar herramientas de forma interactiva.

La multiplicidad de definiciones mostradas permite valorar que las competencias se centran en el hacer en función de una solución demandada. Independientemente de la perspectiva epistemológica, es innegable el hecho que es un elemento que se espera de un profesionista recién egresado, en particular teniendo presente la actual dinámica de rendición de cuentas que ha orillado a la búsqueda de aseguramiento de calidad académica, por supuesto, relacionado con el reconocimiento social de la educación superior (Rodríguez, 2003).

Bajo esta perspectiva y retomando la pregunta inicial *¿Cuáles son las características con las que cuenta un egresado de nivel licenciatura que le permitirán desempeñar su profesión de forma adecuada?* los elementos que conforman las *características con las que cuenta un egresado* incluyen los aspectos éticos descritos (con las dimensiones que en ella se consideran), así como las competencias profesionales que de él se esperan. Al respecto no debe olvidarse que la postura propuesta no considera a estas características como concluyentes ni estables. Como lo considera el filósofo Eduardo Nicol (2013; pág. 108), “el hombre es distinto en cada lugar distinto, porque el lugar es la condición de su modo de *estar* en él”. Bajo esa misma perspectiva, Moreno (2009) establece que las competencias no existen en aislado, siempre dependen del contexto<sup>23</sup>, elemento que deberá tomarse en cuenta al momento de definir dichas competencias.

---

<sup>23</sup> Por lo que el hecho de considerar algunas competencias como *genéricas* o *universales* está en duda (Moreno 2009), lo mismo sucede con propuestas como las formuladas por el proyecto DeSeCo, la cual valdría analizar más a detalle.

La pregunta para definir los aspectos por evaluar rompe con el enfoque tyleriano al momento de plantearse el elemento a futuro (...*que le permitirán...*), ya que considera en su propio planteamiento la imposibilidad de encasillar a un egresado como competente o no competente de forma definitiva. La evaluación propuesta considera entonces que lo evaluado refiere al presente, a aspectos que pueden potencializar el desarrollo profesional de una persona, tanto en el ámbito ético como en el de las competencias, las cuales implican, por supuesto, los conocimientos y habilidades disciplinares.

Teniendo entonces un eje de desarrollo, corresponde a cada ámbito o gremio profesional desarrollar este cuestionamiento y obtener la información respectiva de diversas fuentes de información, en particular, de aquellos *stakeholders* en estrecha relación con el evaluando. La pregunta que nos sirve de eje de desarrollo de la evaluación tiene la característica de ser transversal a las disciplinas, es decir, se puede preguntar *¿cuáles son las características con las que cuenta un egresado de nivel licenciatura que le permitirán desempeñar su profesión de forma adecuada?* a los psicólogos, a los pedagogos, a las diversas ingenierías, a los gastrónomos, economistas y médicos. Las respuestas, por supuesto, podrán incluir elementos que pueden ser identificados como parte de las competencias profesionales, así como elementos éticos. Sin embargo, esto no quiere decir que las respuestas entre disciplinas deban incluir los mismos aspectos e, incluso, que la técnica de recolección de información sea igual en todos los casos.

Para definir de manera particular el marco del evaluando para una disciplina en particular es necesario hacer inicialmente una identificación de los *stakeholders* relevantes para la disciplina en particular. Una perspectiva comprensiva puede tomar como base el diagrama del reloj expuesto en el Capítulo 1, además se requiere hablar de forma constante con los diversos actores, lo que permitirá no solo la identificación de los límites de la evaluación de los recientes profesionistas, sino también debatir sus usos y propósitos. Para una evaluación del egreso de nivel licenciatura se pueden considerar a los siguientes *stakeholders*: alumnos, egresados, directores o coordinadores de carrera, empleadores, padres de familia y docentes. De esta manera, el primer paso para la definición conceptual del evaluando requiere hacer la pregunta *¿cuáles son las características con las que cuenta un egresado de nivel licenciatura que le permitirán desempeñar su profesión de forma adecuada?* a la mayor cantidad posible de *stakeholders* en relación con la disciplina en cuestión. El ejercicio realizado por Ceneval (2009) referido en el capítulo previo marca un antecedente de cómo llevar a cabo esta situación, sin embargo, consideraba tareas profesionales

definidas por un grupo reducido y con una metodología tecnológicamente conveniente, pero paradójicamente limitante en lo que respecta a la amplitud de audiencia de los actores involucrados, al solo centrarse en los egresados de la disciplina. En este sentido, considerar la posibilidad de ampliar al espectro de los agentes participantes favorecerá a la evaluación, ya que desde sus inicios considerará la diversidad de opiniones y perspectivas sobre un mismo fenómeno. Ello favorece la coherencia de la metodología con la epistemología de la cual deviene y alimenta la posibilidad de interpretación de los resultados, como se analizará más adelante.

Epistemológicamente hablando, la delimitación del evaluando representa la construcción de límites conceptuales, de acción, así como éticos que permiten el entendimiento de una identidad de los egresados. En otras palabras, las características con las que cuenta un egresado de nivel licenciatura que le permitirán desempeñar su profesión de forma adecuada será una realidad construida para un entendimiento de aquello que nos interesa observar, ya que las características por observar del egresado no son en sí mismas, sino en función de quién construye sus demarcaciones. Esta construcción teórica se define de forma deductiva y por ende es necesario considerar las fuentes de información o las perspectivas que definen la nube teórica que representa el egreso de una profesión. En el caso de los EGEL, esto se pregunta al Consejo Técnico que, como se comentó, son representantes de algunas IES que ofrecen la disciplina en cuestión. Si bien su punto de vista es válido en tanto que están en contacto directo con el fenómeno de interés, bajo una perspectiva comprensiva resulta necesario considerar la complejidad que representa la diversidad de posturas. Siendo así, es preferible realizar esta construcción teórica desde la opinión de los *stakeholders* en su conjunto (o de los que resulten alcanzables al evaluador ya las condiciones institucionales en las que se encuentre).

Lo que la experiencia de la metodología social a gran escala nos dice es que lo ideal es hacer uso de un muestreo aleatorio (en este caso, sin tener pleno conocimiento del universo total de posibles encuestados), por supuesto, estratificado en función de los diversos *stakeholders*, ello para identificar a las personas a quienes les pediremos responsan nuestra pregunta inicial. El uso de la metodología aleatoria indica bajo los postulados del Teorema Central del Límite, la cual dice que “la suma de un gran número de variables aleatorias independientes tiende a seguir de manera asintótica una distribución normal, siempre que determinadas condiciones queden satisfechas” (Wisniewski y Velasco, 2001; en Alvarado y Batanero, 2006). En otras palabras, el conjunto de observaciones de una muestra tenderá a ser similar a la población de la cual la muestra deviene

mientras mayor sea éste. La metodología inferencial permite tener relativa certeza respecto a que la opinión de las personas a las que se les haga la pregunta será representativa de la población a la cual significan considerando, claro, que ello es un modelo que favorece cierto entendimiento de una aparente objetividad en los datos que se obtengan. Por supuesto, para la construcción de cualquier modelo se requerirá inicialmente contar con ciertas condiciones que, más que representar un reto metodológico, lo son en un sentido práctico. Para acercarse a conocer cuáles pueden ser esas características de los egresados que se pretenden evaluar el eje no podrá ser la metodología aleatoria estratificada (aunque puede considerarse si existe cierta idea o algún registro de los diversos actores con quienes se puede tener cierto acercamiento), sino a quiénes se tiene acceso. Ello esta función de factores contextuales como la consideración si la prueba es nacional. Al ser un proyecto de gran escala implica que se tengan que tener opiniones de actores inmersos en: colegios particulares y privados, zonas urbanas y rurales, escuelas de alta y baja demanda. Adicionalmente se debe considerar la diversidad de género y edades. Las variables que se deben tomar en cuenta deberán ser registradas en el formato de encuesta que se desarrolle para la posterior consideración, ya que ciertas respuestas pueden verse influidas por algunas variables en particular y es eso justamente lo que interesa a la perspectiva comprensiva. Adicionalmente, la planeación debe considerar algunos aspectos de diversidad que sean de interés en particular para el gremio, como la zona geográfica en el caso de los egresados de turismo o la escuela de pensamiento en el caso de los filósofos y los psicólogos, e incluso si la institución se encuentra cerca de una frontera y puerto, para el caso de disciplinas sensibles a ello, como el caso de negocios y comercio internacional. Estos elementos deberán ser tomados en cuenta en el momento de la planeación logística de recolección de información.

La técnica por utilizar que puede recomendarse es, por supuesto, la encuesta. Tradicionalmente la medición de la opinión o actitudes sobre un fenómeno se hace mediante encuestas que utilicen reactivos con escalas Likert (1932). Por supuesto, una escala Likert (como otro tipo de escalas, tal como el diferencial semántico, escalas numéricas o descriptivas) permite una medición precisa de la opinión de una persona sobre un tema en particular; sin embargo, implica la definición del evaluando de los límites o expresiones posibles por parte del encuestado. Al respecto, como se comentó al inicio de este capítulo, Santos Guerra (1988) afirma que la evaluación es un instrumento de poder tal que permite la construcción de una realidad en particular sobre una persona. Siendo así, el uso de una técnica Likert implicaría una definición previa por parte del evaluador de la realidad sobre la cual los encuestados deben responder, sin dar oportunidad a la

inclusión de características de un egresado que el encuestado pueda considerar fuera de lo que la encuesta ya definida le permita. Por tanto, la construcción de la realidad sobre una persona no debe ser decisión del evaluador, sino de la diversidad de perspectivas y contextos que le rodean. Considerando esta limitante, se puede realizar una encuesta abierta, solicitando al encuestado las cinco o diez características más importantes con las que una persona puede egresar de nivel licenciatura para permitirle desempeñar su profesión de forma adecuada.

Como seguramente puede apreciarse, no se presenta una descripción logística detallada del proceso de recolección de información, dado que excede el objeto de la presente propuesta, en particular debido a que la técnica logística para obtener la información deberá ser planeada en función de elementos institucionales como el presupuesto destinado para esta actividad, el personal con el que se cuente (desde el número disponible, así como su perfil), la tecnología disponible (como para realizar la recolección mediante un instrumento en papel y lápiz, o bien usar plataformas Web que faciliten el proceso), las regiones del país en las que se esté interesado, entre otros aspectos. Evidentemente, parte de la viabilidad de esta propuesta metodológica también va relacionada con la posibilidad financiera y de tiempo con la cual se cuente para el desarrollo de esta fase y las subsecuentes. Como es de suponerse, la recolección de información de un alto espectro de *stakeholders* bajo las características que se proponen no resulta una implementación sencilla o barata, debido a los gastos en personal y viáticos que ello requiere, lo cual debe tenerse presente.

Por supuesto, el análisis de la información que se obtenga debe ser cuidadoso, debido a que ello representa la definición de lo que se pretende evaluar; es decir, con este paso comienza el establecimiento de la realidad en particular que se observará, que a su vez permitirá el desarrollo de estándares con los cuales cada egresado evaluado será comparado y retroalimentado. Considerando la responsabilidad que este paso implica debe realizarse con detenimiento y cuidado.

Posteriormente, para efectos del análisis de la información puede hacerse uso de una metodología de análisis de contenido, en particular de corte descriptivo (Pinuel, 2002). Este tipo de análisis consta en recopilar las observaciones de las personas encuestadas con el fin de identificar y catalogar la “realidad empírica de los textos o documentos, mediante la definición de categorías o clases de sus elementos” (Pinuel, 2002, pág. 9). No se trata solo de hacer una simple clasificación de las opiniones de las personas, sino que se debe hacer de forma cuidadosa, tomando en cuenta

los contextos de donde se emiten las opiniones. Esta metodología permite hacer una clasificación que favorece la cuantificación de las opiniones, así como la descripción cualitativa de lo que implica cada una de ellas en función del contexto bajo el cual se emiten, delimitado por las variables que hayan sido descritas en los formatos de recolección de información.

Como resultado, se espera una aproximación a una realidad en particular, delimitada no por la perspectiva del evaluador (solo mediada por él), sino por una diversidad de posturas hacia un mismo fenómeno considerando a un grupo de *stakeholders* heterogéneos. Por ende, el producto final deberá ser entonces el perfil del egresado que se pretende evaluar, el cual estará definido completamente por el conjunto de *stakeholders* de la disciplina en particular. Dicho perfil deberá describir de forma detallada las categorías obtenidas del análisis de contenido, las cuales fungirán como las áreas por evaluar, dentro de las cuales pueden estar incluidos elementos éticos propios del quehacer profesional en cuestión; por supuesto, cada una de las áreas requerirá del establecimiento de los elementos que la componen mediante una definición. Por ejemplo, si una de las áreas fuera “Realizar diagnósticos” para el caso de una formación médica, entonces se tendría que definir los aspectos que conforman el desarrollo de diagnósticos, inmerso en ello, por supuesto, los elementos de competencia y principios éticos que deben definir este quehacer. La definición de los elementos que lo conforman puede ser establecida mediante elementos bibliográficos, o bien, mediante la revisión de los planes de estudios de las instituciones participantes. En ocasiones las propias respuestas de los encuestados pueden dar orientaciones sobre los elementos que conforman cada categoría, sin embargo, esto requerirá ser consultado con referentes más formales de la disciplina, como los ejemplos mencionados.

Bajo esta metodología se puede vislumbrar que el proceso para definir al evaluando deberá establecerse mediante una planeación acorde con la disciplina en particular, considerando también los recursos con los que se cuenta para este trabajo, siempre teniendo cuidado en su conformación, ya que este es el proceso más delicado del cual dependerá todo el proyecto de evaluación. Ello implica no olvidar el tiempo necesario para su realización.

#### *Objetivo de la evaluación*

Retomando la discusión inicial que orilló a la definición de los aspectos necesarios para delimitar al evaluando, ahora que se ha identificado el objeto de la evaluación, el objetivo puede verse con mayor claridad. Recordando los postulados del enfoque comprensivo, orientados hacia la

consideración de los elementos sociales en lo observado (Stake, 2006), el objetivo de la evaluación del egreso de nivel licenciatura deberá estar enfocado a la retroalimentación de los programas de estudios que participen en el proyecto más que en reportar niveles de rendimiento que caracteriza a la evaluación que Stake denomina *basada en estándares*, así como en que los egresados participantes conozcan el estatus de su preparación en función de los límites conceptuales y de la praxis de su disciplina enmarcada en la evaluación. No se sugiere su uso con fines de titulación, como requisito de egreso o con la intención de obtener medallas, ya que, como se analizó en el Capítulo 3, este tipo de prácticas no favorecen los fines formativos de una evaluación y pueden orillar al desarrollo de fenómenos nocivos como la enseñanza para la prueba y que todos los esfuerzos para llevar a cabo una evaluación de tanta complejidad se reduzcan a una prueba o un simple ranqueo de resultados.

Por supuesto, dada la necesidad del contexto actual centrado en políticas de rendición de cuentas, así como los antecedentes de las IES nacionales que dieron origen al EGEL, es necesario que este ejercicio se desarrolle a gran escala, ya que de esa manera se pueden dar resultados bajo un estándar de construcción común entre instituciones que favorezca las razones que en primer lugar dieron pauta a la generación de las evaluaciones del egreso en México. Por supuesto, no se debe olvidar que a pesar de plantear una evaluación bajo un objetivo formativo que facilitaría el cumplimiento de las metas ideales que dieron pauta al EGEL, se tiene que considerar la necesidad administrativa y política que favoreció su consolidación en nuestro país. En ese sentido, el desarrollo de una evaluación a gran escala con un objetivo esencialmente formativo no debe descuidar la necesidad administrativa que soluciona la metodología a gran escala. Por tanto, pensar en la posibilidad de particularizar de forma absoluta de estos ejercicios de evaluación nos llevaría a la no consideración del ejercicio inicial que dio vida a esta práctica educativa en nuestro país: la mejora de los programas de educación superior. Además, las prácticas de evaluación del egreso se reducirían a las actividades con fines de titulación que se realizan de forma cotidiana dentro de cada institución, las cuales son profundamente valiosas y alimentadoras, pero que no favorecen a la búsqueda de la mejora de la educación intra e inter institucional perseguida en nuestro país.

#### *Construcción de instrumentos*

Previo a la descripción de la propuesta para el desarrollo de instrumentos, vale considerar que bajo una perspectiva comprensiva la participación de diversos agentes en el desarrollo de

cualquier actividad es un elemento primordial. Por ello, durante este subapartado se valora la sugerencia de participación de diversos agentes además de la propia instancia evaluadora (o la agencia o departamento encargado del desarrollo del proyecto).

La primera fase metodológica para el desarrollo de la evaluación de forma comprensiva implica la definición de los límites de lo que se pretende evaluar. Es una metodología comúnmente aceptada en el desarrollo de la investigación y la evaluación educativa, como se ha comentado, requiere la construcción teórica de lo que se observará, lo cual impondrá dentro de sí mismo los límites sobre lo que se darán resultados. Si esos límites son definidos por los *stakeholders* que están en contacto directo con el evaluando y que lo conocen a fondo, nos permitirá establecer una postura sobre el evaluando cercana a la diversidad de perspectivas y realidades que en su entretrejo conformarán una postura en particular.

Una vez definidos esos límites, es necesario establecer las estrategias de recolección de información que permitirán conocer al evaluando con fines de retroalimentación. Por supuesto, el instrumento tradicionalmente utilizado son las pruebas de rendimiento, dado el supuesto que la educación solo se refleja en resultados de aprendizaje, y si bien habrá que considerar que la definición de los elementos por evaluar dependerá mucho de las respuestas que los *stakeholders* den a la pregunta inicial, se pueden considerar diversos instrumentos que en su conjunto complementen la visión sobre el egresado de nivel licenciatura.

El instrumento y la fuente de información por utilizar dependerán en gran medida de la naturaleza de la categoría de las características del egresado y los elementos que la determinan. Si bien el objeto de este trabajo no es presentar los pasos de diseño de todos los tipos de instrumentos que se pueden utilizar para la recolección de información, cabe la descripción de los elementos característicos de los instrumentos que pueden implementarse en función de las diversas posibilidades por enfrentar.

#### a) *Prueba*

Como se observó en el Capítulo 2, las pruebas de rendimiento o de logro educativo son los instrumentos más usados en este tipo de evaluaciones. Aún bajo una perspectiva comprensiva, es válido el uso de instrumentos estandarizados, aunque son considerados como un elemento adicional a un conjunto de formas de recolección de información (Stake, 2006). Ahora, también se debe considerar que en la actualidad y bajo el contexto que actualmente se enmarca la educación

superior y la evaluación en general, pensar en una prueba de rendimiento o de conocimientos resulta insuficiente y realmente limitado, sobre todo en los grandes propósitos que la evaluación del egreso tiene para la política educativa de cada una de las IES dentro del territorio nacional. Por supuesto, como se ha comentado, la medición de rasgos psicológicos como el aprendizaje o la opinión es necesaria, aunque vale no considerar los resultados como un elemento definitorio o una *verdad absoluta*.

La tendencia educativa actual, justo como se comentó en el apartado previo respecto a la definición del evaluando, es la de considerar a las competencias, en función que los planes de estudio tienden a estar centrados la incorporación de metodologías cercanas a la realidad profesional (Fernández, 2010). Por la forma en que se propone realizar una encuesta a diversos actores, los conocimientos, habilidades y actitudes deberán ser evaluadas entre los egresados bajo una perspectiva metodológica mucho más amplia que la que actualmente se desarrolla en México.

La bibliografía sobre la evaluación de competencias indica que ésta tiene particularmente una perspectiva videográfica y no fotográfica (Tardif, 2006, en Fernández, 2010), es decir, debe ser una práctica que valore los elementos de proceso, más que una cuestión sumativa. Esta postura afirma que la evaluación de las competencias debe realizarse de forma continua, sin embargo, es algo que no necesariamente puede desarrollarse bajo el objetivo de la evaluación planteado en esta propuesta, dada la idea central en el egreso; aunque el hecho de que sea centrada en el egreso no significa que no sea formativa, dado que a pesar de ser al final de un ciclo educativo formal la idea de la evaluación es realizar una retroalimentación a los planes de estudio de las IES, así como para los recién egresados. Para llevar a cabo esta tarea, la medición de las competencias debe basarse en el acceso a fuentes múltiples y variadas de información (Fernández, 2010).

La bibliografía propone diversas técnicas e instrumentos para evaluar aspectos complejos como lo son las competencias, al respecto, Vila y Poblete (2011) afirman que algunos principios que se deben de considerar para llevar a cabo una evaluación de competencias son: evaluarlas de forma integral (no como elementos aislados); considerar instrumentos o técnicas con mayor valor añadido para garantizar la recolección de información; no se debe olvidar que los conocimientos no son algo separado de las actitudes y valores; requiere además de la determinación de niveles de dominio; debe considerar la experiencia personal (como síntesis del conocimiento). Con base en ello, estos autores presentan el siguiente cuadro resumen que incluye las posibles técnicas e

instrumentos por utilizar para poder observar los diversos elementos que conforman las competencias.

**Tabla 5.1 Técnicas e instrumentos por utilizar para la evaluación de competencias.**

Elementos de competencias por evaluar	Valores					x	X	x	x	x
	Actitudes, intereses					x	X	x	x	x
	Técnicas				x	x	X	x	x	x
	Normas, procedimientos		x	x	x	x	X	x	x	x
	Habilidades, destrezas		x	x	x	x	X	x	x	x
	Conocimientos	x	x	x	x	x	X	x	x	x
	Prueba objetiva	Examen escrito	Problema algorítmico	Examen verbal	Informe (proyecto, caso)	Protocolos de observación	Portafolio	Entrevista	Simulación de situaciones	
<b>Técnicas e instrumentos</b>										

Fuente: Vila y Poblete (2011).

La propuesta de Vila y Poblete (2011) se enfoca en la presentación de retos o actividades que permitan la observación directa de los elementos de las competencias. Como se observa, la llamada *prueba objetiva* se centra particularmente en conocimientos y, si bien en algunos casos se pueden medir aspectos de aplicación de forma limitada, vale considerar aspectos como el portafolio, entrevista, informes y protocolos de observación (de actividades prácticas) como más apropiados para evaluación de competencias, ya que favorecen la observación de conocimientos, habilidades y elementos de índole ético. Por supuesto, como menciona Fernández (2010), se debe cuidar que las estrategias utilizadas sean semejantes a las que se desarrollan en el ejercicio de la profesión, cercanas a las condiciones de contexto de trabajo al igual que el contexto social; esto implica, por supuesto, la observación de diversos criterios. Por tanto, la evaluación de estos aspectos debe considerar una diversidad de instrumentos (Cano, 2008). Al respecto, (Bolívar, 2008, en Cano, 2008; pág. 12) menciona que:

“La mejor manera de evaluar competencias es poner al sujeto ante una tarea compleja, para ver cómo consigue comprenderla y conseguir resolverla movilizándolo conocimientos. Los instrumentos de evaluación empleados no pueden limitarse a pruebas para ver el grado de dominio de contenidos u objetivos sino proponer unas situaciones complejas, pertenecientes a la familia de las situaciones definida por la competencia, que necesitará por parte del alumno, asimismo, una producción compleja para resolver la situación, puesto que necesita conocimiento, actitudes, pensamiento metacognitivo y estratégico”.

Estas acepciones representan problemas técnicos en dos vías para la presente propuesta. El primero de ellos refiere a la metodología a gran escala que define el proyecto. La denominación *a gran escala* implica no un objetivo, sino una definición operativa y metodológica de su quehacer.

Si se pretende dar resultados a los estudiantes, pero además a las IES, entonces no es viable el uso de algunos instrumentos propuestos por Villa y Poblete (2011) por las siguientes razones. El *Examen verbal*, los *Protocolos de observación*, la *Entrevista* y la *Simulación de situaciones* implican la revisión a profundidad de las habilidades, intereses, valores y uso de técnicas por parte de uno o varios jueces de todos los egresados participantes. Si bien se pueden desarrollar escalas o rúbricas que permitan la codificación de las observaciones en todos estos casos, resulta operativamente y económicamente inviable considerar la asignación de más de un juez que valore lo observado y asigne los criterios correspondientes. Justo ahí reside la ventaja del uso de pruebas objetivas en la evaluación a gran escala, ya que, como se comentó en el Capítulo 2, la codificación de las respuestas a los reactivos de este tipo de instrumentos (que son mayoritariamente reactivos de opción múltiple) resulta práctica y sencilla. Sin embargo, dada la complejidad de una evaluación del egreso bajo una perspectiva comprensiva, el uso de este tipo de instrumentos debe verse reducido de la mayor forma posible.

El segundo problema se refiere a los elementos éticos que son propios del desempeño profesional, los cuales, sin ser medidos de una forma tradicional en las pruebas de gran escala mediante reactivos de opción múltiple (ya sean dicotómicos o politómicos), puede orillar a la reducción de los elementos éticos a un simple *deber ser*; es decir, las respuestas *correctas* de este tipo de reactivos recaerán siempre en acciones de lo que moralmente se *tienen* que realizar, pero este formato no permitirá medir los elementos que los estudiantes consideran en su toma de decisiones profesionales.

Considerando que lo que se espera del perfil del evaluando definido son elementos de competencias y algunos posibles aspectos éticos, se sugiere el uso de un instrumento que considere la medición de habilidades, conocimientos y formas de enfrentar los problemas profesionales. Es factible hacer uso de reactivos de opción múltiple para aquellos contenidos netamente disciplinares, pero se debe considerar sobre todo la posibilidad de reactivos abiertos para el caso de habilidades y acciones éticas.

Para el desarrollo de reactivos se debe tomar como principal insumo la categorización de los elementos resultantes de la encuesta a los *stakeholders*. Dicha información deberá redactarse considerando la pregunta *¿qué se espera medir de dicha categoría?* Esta pregunta es clave en la redacción del esquema de conocimientos, habilidades acciones éticas que permitirán definir los reactivos por utilizar en el instrumento. Si bien tradicionalmente se cuida una redacción

mayoritariamente operacionalizada, es decir, que refleje de forma fehaciente el elemento por medir de manera explícita, también es posible considerar una definición más abarcativa. Como ejemplo, tomemos una definición operacional obtenida de las especificaciones que el Ceneval ha hecho públicas en su texto *Manual técnico* (Ceneval, 2013b), la cual se centra en una perspectiva atomista y se presenta el ejercicio de la definición de los elementos por evaluar bajo una forma más comprensiva, la cual evite el aislamiento de habilidades y fomente la inclusión de habilidades, conocimientos y aspectos éticos inmersos en las prácticas profesionales:

**Tabla 5.1 Ejemplo de redacción de elementos por evaluar.**

<b>Definición operacional EGEL en Turismo</b>	<b>Definición de elemento por evaluar bajo una perspectiva comprensiva</b>
A partir de la descripción de la situación de mercado de un producto o servicio turístico, el sustentante determina la o las técnicas para pronosticar la demanda total.	El egresado decide la o las técnicas para pronosticar la demanda total de un producto o servicio turístico, considerando en su elección el contexto del producto o servicio.

Fuente: Elaboración propia, considerando información de Ceneval (2013b).

En la definición utilizada por el EGEL en Turismo, la redacción de la definición operacional está diseñada con la finalidad de mantener poca amplitud de decisión para quien elabora los reactivos, dando lugar a reactivos simplistas con una breve redacción de contexto y cuatro opciones de respuesta para que el egresado seleccione la que permite calcular la demanda total. A continuación el ejemplo de reactivo que se desprende de esta definición operacional mostrado en el mismo documento:

**Tabla 5.2 Reactivo ejemplo del EGEL en Turismo.**

Un hotel cuenta con los siguientes datos sobre el mercado nacional e internacional:

- Existen 4,500,000 turistas internos y 12,800,000 turistas receptivos
- Los turistas internos adquieren dos noches en promedio de hospedaje y los turistas receptivos siete noches en promedio
- El costo por noche es de 57 USD para interno y 48 USD para el receptivo

Con el fin de pronosticar la demanda total del mercado nacional de producto, es necesario utilizar:

- A) las variables que determinan la demanda  
Incorrecta. Argumentación<sup>24</sup>: En esta técnica se utilizan variables como aumento de población o el ingreso per cápita, que es información con la que el hotel no cuenta.
- B) el método de composición del mercado  
Correcta. Argumentación: Requiere los datos de los compradores y sus compras para determinar la estructura del mercado al cual se dirige el hotel
- C) el cálculo de los semipromedios  
Incorrecta. Argumentación: Implica el uso de estimaciones lineales, que no es el insumo bajo el cual el hotel está determinando su demanda
- D) el índice de factores del mercado  
Incorrecta. Argumentación: Se identifican los factores del mercado correlacionados con el potencial del mercado, que son insumos que no está utilizando el hotel en este caso en particular.

Fuente: Ceneval (2013b).

Como se puede apreciar en el reactivo ejemplo, es un claro espejo de la definición operacional, lo cual asegura consistencia en la medición; sin embargo, lo que se alcanza a medir es limitado, ya que no están en juego actitudes o aspectos éticos que los profesionistas ponen en juego al decidir una técnica en particular. Para su resolución basta con conocer los conceptos e hilarlos a las condiciones del contexto que se presenta en la base del reactivo. Ahora bien, a manera de comparación a continuación se presenta un ejemplo de reactivo a partir del elemento por evaluar definido de una manera comprensiva.

---

<sup>24</sup> La "Argumentación" es un elemento que conforma a cada reactivo que se incorpora al EGEL, sin embargo, no se muestra a los egresados; solo se desarrolla con fines de dar solidez académica a cada opción de respuesta y así asegurar su plausibilidad.

**Tabla 5.3 Reactivo ejemplo bajo la perspectiva comprensiva.**

A mediados de año, un hotel *boutique* de la zona centro de Guadalajara, Jalisco, se ve en la necesidad de pronosticar la demanda total de sus servicios de alojamiento, ello con la finalidad de planear su presupuesto para el siguiente año. En años previos han observado que la demanda total se ve influida fuertemente por los *puentes* y fines de semana largos, así como por el festival anual gastronómico que se dan en la región, dado el interés del Gobierno Municipal de fomentar el turismo gastronómico. En particular, estos elementos influyen en un incremento de demanda total de hasta 70% respecto a la temporada baja.

Considerando el objetivo y las variables descritas, ¿cuál de los siguientes métodos permitirá hacer la estimación más precisa de la demanda total del alojamiento?

- A) Análisis de la composición del mercado
- B) Cálculo del índice de factores del mercado

Argumente su respuesta:

---

---

Fuente: Elaboración propia.

El ejemplo mostrado en la tabla 5.3, guarda estrecha coincidencia con la definición del *Elemento por evaluar*, aunque el esquema de respuesta sufre modificaciones que permiten hacer un análisis a mayor profundidad de las habilidades y decisiones del egresado. Las dos opciones de respuestas pueden ser correctas, ya que no hay elementos en la base del reactivo que condicionen una decisión u otra (como sí sucede en el ejemplo del EGEL en Turismo), sino que lo interesante es analizar los factores de decisión del egresado. Para el ejemplo en particular se incluyen dos opciones únicamente, debido a que para el caso presentado son dos decisiones plausibles, aunque ello puede variar. Este análisis permitirá conocer lo estipulado en el *Elemento por evaluar*: "... decide la o las técnicas para pronosticar la demanda total de un producto o servicio turístico, considerando en su elección el contexto del producto o servicio". Por supuesto, este tipo de reactivos deben ir acompañados de una rúbrica que permita la codificación de las diversas respuestas.

Las rúbricas son "escalas de evaluación que incluyen niveles de dominio progresivo acerca de un determinado conocimiento" (Romero, 2012; pág. 43). Son un tipo de escalas de valoración que permiten comparar una respuesta u observación con criterios en particular que son graduados conforme a un criterio en particular (normalmente referenciados a la complejidad de ejecución). Existen dos tipos principales de rúbricas: las holísticas y las analíticas. Las holísticas requiere la codificación de un proceso o producto como un todo, mientras que las analíticas se codifican de

forma separada los elementos que conforman un proceso o un producto y al final se suman los valores de cada elemento para obtener un puntaje total (Moskal, 2000; Nitko, 2001; ambos en Mertler, 2001). Ambos tipos de rúbricas son altamente utilizados, pero en particular las analíticas son las que de forma más frecuente se utilizan dentro del aula, ya que favorecen el conocimiento preciso de todos los conocimientos y habilidades involucrados en un proceso, lo cual las hace altamente recomendables para objetivos formativos. Para esta propuesta se sugiere el uso de rúbricas holísticas, tal como se hace en otros ejercicios a gran escala, como el caso de PISA y LLECE. En ambos proyectos las rúbricas holísticas permiten desarrollar un proceso de codificación de respuestas a reactivos abiertos relativamente sencillo y rápido. Más adelante se describe el proceso de codificación propuesto mediante rúbricas holísticas.

Hasta ahora, la metodología propuesta para el desarrollo del instrumento puede resumirse mediante el siguiente esquema, considerando ya una definición previa del evaluando:

**Gráfica 5.1 Resumen del desarrollo de la prueba.**



Fuente: Elaboración propia.

Los cuatro grandes rubros de esta propuesta son particularmente acciones metodológicas que siguen una perspectiva de trabajo en particular, pero en cada una de ellas se requiere la intervención de los *stakeholders*, o bien, de representantes disciplinares. Estos representantes, quienes no se habían mencionado previamente en esta propuesta, son resultantes de la experiencia de los CT y CA que se da en los EGEL. En este caso, se definen como *representantes disciplinares*, ya que no necesariamente deben constituir un órgano fijo como lo es regularmente el equipo metodológico de la instancia evaluadora. Esta figura se constituye por personas representantes de la disciplina, más que a una escuela o institución. Estas personas deben ser las encargadas de ayudar al equipo metodológico a definir las categorías, a redactar los elementos por evaluar y a la redacción de reactivos. Por supuesto, deben ser personas que hayan tenido una

formación en la disciplina en cuestión, así como experiencia laboral y/o docente en el área. Corresponde a la instancia evaluadora la consideración de los aspectos de contratación en función de presupuesto y tiempos para la ejecución del proyecto. Lo interesante en esta propuesta resulta de resaltar la relevancia de contar con la mayor diversidad de opiniones posible en la definición del evaluando y con expertos disciplinares en la orientación propia de las particularidades de la prueba. Evidentemente, en este proceso, al igual que en el resto de la propuesta, los recursos financieros y de tiempo, así como la participación de diversos agentes tendrán que desarrollarse considerando las necesidades prácticas en las cuales esté inmerso el proyecto.

Como elemento resultante, la prueba debe considerar una diversidad de categorías de los egresados, tomando una definición de todos los elementos por evaluar, los cuales a su vez fungirán como insumos en el desarrollo de reactivos metodológicamente complejos que vayan acompañados de rúbricas de codificación holísticas. El desarrollo de reactivos deberá considerar, en la medida de lo posible, la participación de representantes disciplinares o especialistas del área bajo la orientación metodológica de la instancia evaluadora. En lo que respecta a las rúbricas, más adelante se detalla su desarrollo.

El número de reactivos deberá estar en función del universo de elementos por evaluar, aunque la cantidad no deberá ser tal que supere la capacidad de concentración e integridad de los egresados. La experiencia de evaluaciones nacionales e internacionales aplicadas en México nos dice que el interés por mantener una validez de contenido de la prueba, es decir, de poder medir todo el constructo posible, ha orillado a que los EGEL duren 4 horas (en el caso de las aplicaciones piloto, éstas pueden durar hasta 8 o 12 horas, por supuesto abarcando descansos o hasta dos días en su ejecución), o bien, 4 horas en pruebas matriciales con una enorme cantidad de reactivos (tal el caso de PISA). El desarrollo de reactivos debe cuidar no tener un impacto de tal magnitud en el cansancio de los egresados, sobre todo por el hecho de que no está considerada como una prueba matricial, en virtud del interés de emitir una retroalimentación a los egresados participantes. Si bien la llamada validez de contenido es relevante, hay que considerar que la utilización de reactivos abiertos, como el mostrado en la Tabla 5.3, permite una valoración de mayor profundidad respecto a las decisiones profesionales de los egresados, en ese sentido, una propuesta como la que se presenta, da un mayor valor a la profundidad de la información que se puede obtener para dar una retroalimentación con sólidas bases empíricas, que a un número importante de reactivos.

## b) Cuestionarios

Regularmente una propuesta de evaluación a gran escala considera la aplicación de cuestionarios de contexto, aunque en raras ocasiones se analizan y se emiten consideraciones al respecto. Se pueden valorar algunos ejemplos publicados. Las publicaciones de la OCDE respecto al proyecto PISA incluyen diversos volúmenes que complementan los resultados cognitivos. Por ejemplo, para el caso de la aplicación PISA 2012 se consideraron análisis que permiten contextualizar los resultados, tal como el volumen II, centrado en la equidad, el volumen III enfocado en el compromiso y la autopercepción de los estudiantes, el volumen IV que analiza los recursos y prácticas escolares<sup>25</sup>. Estos textos permiten considerar diversos aspectos que, bajo la perspectiva de la OCDE, *complementan* la medición cognitiva. En contraste, el Ceneval publica sus informes anuales, los cuales conjuntan los datos de los sustentantes que aplican el EGEL durante un año seguido. El informe muestra de forma general, la correlación estadística entre los resultados en la prueba y algunas variables que son consideradas de contexto, como el género del estudiante, su promedio de licenciatura, si su escuela era pública o privada o si recibía beca durante sus estudios. La información de estos informes anuales del Ceneval es sin duda interesante, sin embargo, no representan una verdadera contextualización de resultados, ya que no son elementos de influencia para la determinación de una retroalimentación y, más aún, no se consideran para efectos de comprensión de los resultados.

Si una propuesta de evaluación se considera *comprehensiva*, por supuesto debe tomar en cuenta la diversidad que representa una respuesta dada a un reactivo. Si bien el elemento a gran escala impide hacer un análisis a profundidad de cada respuesta a cada reactivo, es necesario esbozar un contexto en función de los resultados de la prueba, no porque ésta última sea el eje mismo de la evaluación, como lo considera PISA, o porque sea complementaria a los resultados, como lo exponen González-Montesinos y Backhoff (2010), sino porque es el contexto lo que define y explica una situación en particular de los egresados.

Son llamados *cuestionarios* y no *pruebas* debido a que los cuestionarios no tienen la intención de *probar* si la persona cuenta con un elemento de interés de los *stakeholders*, sino porque *cuestionan* sobre sus condiciones. En ese sentido, su constitución no debe estar en función de un desempeño, sino del contexto o realidad circundante del egresado. Este tipo de cuestionarios son

---

<sup>25</sup> Todos los volúmenes se pueden consultar en: <http://www.oecd.org>

procedimientos que permiten entonces recoger información sobre las personas, grupos y escuelas a las que se les aplican otros instrumentos (Jornet, González y Perales, 2012). De hecho, en estricto sentido, no se trata de un instrumento en aislado, sino de un conjunto de instrumentos dirigidos a diferentes audiencias.

De manera similar a las pruebas, el diseño de cuestionarios implica la definición de variables en función del objetivo de investigación o evaluación (Jiménez, García y González, 2006). La definición de variables deberá considerar su nivel y escala de medición, de tal forma que se pueda determinar si el instrumento será abierto o cerrado. Normalmente este tipo de instrumentos se desarrollan centrándose en los factores o variables que la bibliografía muestra como relacionadas con la explicación del desempeño (Jornet, González y Perales, 2012; OCDE, 2015a).

Al respecto, la experiencia internacional de la OCDE dentro del proyecto PISA resulta valiosa. Como se comentó previamente, PISA rescata mucha información de diversos actores y desarrolla marcos de explicación de sus resultados en función de la información proporcionada en sus diversos instrumentos. Se preguntan sobre los antecedentes de los estudiantes (educación en la infancia, estatus social, económico y cultural); enseñanza y aprendizaje (aptitudes y prácticas docentes, clima escolar y tiempo destinado al estudio), así como políticas escolares y gobernanza (organización curricular, gestión escolar, involucramiento de los padres, así como el uso de evaluaciones).

Los cuestionarios en esta propuesta siguen una lógica comprensiva, es decir, no solo la de complementar información, sino de encuadrar las características de los egresados en función de un contexto. Al respecto, no se encontró información relacionada con el análisis de los elementos de contexto relacionados con el egreso. Si bien hay bibliografía relacionada con un egreso *exitoso* (como las horas de dedicación al estudio, promedio de nivel medio superior, entre otras variables), la presente propuesta no se centra solo en la identificación de egresados *exitosos*, sino en la caracterización del momento social que define a los egresados de una disciplina en particular. Para cada disciplina hay variables que adquieren mayor importancia que otras. Por ejemplo, la experiencia laboral previa al egreso puede ser más valiosa para un estudiante de leyes que para uno de letras; o bien, los antecedentes escolares pueden ser más relevantes para filosofía que para gastronomía, dada la importancia en el desarrollo de habilidades de razonamiento *versus* el desarrollo de habilidades manuales. Como consecuencia, corresponde a cada disciplina, mediante un análisis detallado de los planes de estudio, así como de la perspectiva de los representantes

disciplinarios, identificar qué variables de contexto son relevantes para enmarcar las características de los egresados. En este análisis se sugiere la consideración de los siguientes rubros contextuales, los cuales pueden representar en primer momento una guía de análisis sobre la cual cada evaluación puede añadir o modificar elementos.

**Tabla 5.4. Ejemplo de rubros y variables por considerar en el desarrollo de cuestionarios.**

Rubros de análisis	1	2	3	4	5	6
	Trayectoria académica del estudiante	Antecedentes personales y sociales del estudiante	Experiencia al cursar la disciplina y expectativas laborales	Perspectivas sobre la preparación de los egresados hacia su futuro laboral	Condiciones escolares	Ventajas y limitantes del plan de estudios
Variables por considerar	<p>Condicionantes que le orillaron a tomar esa disciplina</p> <p>Años de escolaridad</p> <p>Años requeridos para el egreso</p> <p>Promedio obtenido</p> <p>Habilidades que más esfuerzo costó obtener</p>	<p>Apoyo de la familia para el desarrollo de estudios</p> <p>Redes de apoyo</p> <p>Objetivo de estudios profesionales</p> <p>Estudios de los padres</p>	<p>Principales habilidades requeridas para concluir con éxitos los estudios</p> <p>Expectativas académicas</p> <p>Expectativas laborales</p> <p>Principales fortalezas desarrolladas</p> <p>Principales habilidades no desarrolladas</p> <p>Principales retos durante la formación</p>	<p>Principales fortalezas de los egresados en cuestión</p> <p>Principales debilidades de los egresados en cuestión</p> <p>Posibilidades de colocación laboral</p> <p>Pertinencia actual del plan de estudios</p> <p>Suficiencia del plan de estudios</p>	<p>Nivel de preparación didáctica de los docentes</p> <p>Utilidad del servicio social y prácticas profesionales</p> <p>Efectividad del servicio de tutoría</p> <p>Efectividad de movilidad académica</p> <p>Condiciones de infraestructura relacionada con la formación en la disciplina</p> <p>Relación con empleadores</p>	<p>Pertinencia actual del plan de estudios</p> <p>Suficiencia del plan de estudios</p> <p>Efectividad del plan de estudios</p>
Fuente de información	<b>Estudiante</b>			<b>Docentes</b>	<b>Director / Jefe de carrera</b>	

Fuente: elaboración propia.

Como se observa en el ejemplo, es necesario considerar diversas fuentes de información, desde el estudiante mismo, así como algunos docentes (que pueden ser seleccionados de manera aleatoria,

o bien, en función de algunas asignaturas clave para la disciplina en cuestión –aunque ello puede representar un problema de equivalencia entre planes de estudio–) y el director o jefe de carrera. Las variables que conforman cada rubro deben considerar una perspectiva comprensiva y la opinión de los agentes involucrados. La intención de los cuestionarios será entonces la búsqueda de la comprensión del contexto de los estudiantes, los problemas y retos a los que tuvieron que enfrentar y las condiciones bajo las cuales desarrollaron sus estudios. No se pretende encontrar elementos contextuales que permitan emitir resultados diferenciados para cada situación, ya que ello dificultaría la interpretación respecto a lo esperado por los egresados de una misma disciplina<sup>26</sup>, la intención es comprender a cabalidad las condiciones contextuales de los egresados, lo cual también implica una influencia en su formación como profesionistas. Los cuestionarios permitirán retroalimentar de forma profunda a las instituciones participantes en la evaluación, dado que darán la oportunidad de enmarcar una retroalimentación particularizando las condiciones de cada institución y, si así se desea, de cada estudiante participante. Corresponde entonces a cada gremio disciplinar la reflexión sobre qué rubros y variables pueden incluirse en la definición contextual de los recién egresados mediante una fuerte guía metodológica de la instancia evaluadora.

El desarrollo de los instrumentos será consecuencia de la definición de los rubros y variables que realicen los representantes disciplinares, así como la instancia evaluadora. Como lo comentan Jornet, González y Perales (2012) y Jiménez, García y González (2006), el siguiente paso corresponderá a la definición de indicadores resultantes de las variables, que son la redacción de las variables que permiten su medición en forma de reactivos. Por supuesto, al igual que con la prueba, se deberá cuidar la consideración de reactivos de respuesta construida siempre y cuando esto sea necesario, considerando, por supuesto, su factibilidad en términos de presupuesto y de utilidad de análisis, aunque comúnmente en este tipo de instrumentos pueden usarse escalas dicotómicas y politómicas.

Con el desarrollo de los cuestionarios concluye la fase metodológica del diseño de instrumentos. Cabe resaltar que la metodología propuesta requiere considerar forzosamente las condiciones institucionales bajo las cuales se desarrolle, desde cantidad de personal, recurso presupuestal y tiempos de ejecución. Por supuesto, es necesario tomar en cuenta que esta fase deberá

---

<sup>26</sup> A menos que así se requiera, como en disciplinas cuya enseñanza varía fuertemente en función del contexto, como en el caso de Derecho o Comercio Internacional.

desarrollarse lo más apegado al objetivo y a un enfoque comprensivo, considerando como eje la diversidad de perspectivas, opiniones y contextos, mismos que deben alimentar el proyecto; de tal forma que los instrumentos que se desarrollen no funjan como punto nodal de la evaluación, sino como una consecuencia de la conjunción de la diversidad de perspectivas sobre un fenómeno en particular.

Tampoco se debe olvidar que es deseable los instrumentos deben pasar por una fase de piloteo, en la cual se pruebe la manera en que todas las preguntas están redactadas, la posibilidad de incluir opciones de respuesta diferentes a las primeramente estipuladas, así como la factibilidad de tiempos de aplicación. Por supuesto, ello depende en gran medida de los tiempos disponibles para el proyecto, el personal participante y el presupuesto destinado para ello. En condiciones no óptimas los instrumentos pueden probarse mediante el juicio de expertos disciplinares, o bien, *stakeholders* que participen en la valoración de la articulación y redacción adecuada de los instrumentos. Aunque la consideración metodológica basada en la participación de diversos grupos de trabajo debe favorecer a la consideración de la complejidad del fenómeno del egreso de la licenciatura mediante los instrumentos que se elaboren. De esta forma, el uso de pruebas y cuestionarios favorecerá una comprensión profunda de las condiciones de egreso de los estudiantes, así como los elementos institucionales que influyeron en ello.

### Proceso de recopilación de información

El elemento operativo que forma parte de cualquier evaluación, y en particular aquellas con un enfoque a gran escala, representa un aspecto de gran importancia y cuidado. Como se comentó en el capítulo previo, ha habido una búsqueda incesante de aspectos sistemáticos con el fin de no poner en riesgo la validez de los instrumentos aplicados. Por supuesto, es necesario establecer un esquema de aplicación que facilite la recopilación de información, ello en función de los instrumentos mismos además de la disponibilidad de espacios, tiempos previos a la aplicación (como el dedicado a la impresión de materiales y a la capacitación de aplicadores), así como también de los recursos requeridos y disponibles.

La experiencia de Ceneval en los EGEL nos dice que es viable considerar diversos tiempos de aplicación al año, así como el uso de sistemas informáticos para la aplicación. Para el caso de una evaluación comprensiva, es deseable considerar únicamente una evaluación al año, ello permitirá la dedicación adecuada de tiempo y recursos para una aplicación de grandes magnitudes y, en

particular, dedicar lo suficiente a la codificación de respuestas abiertas y desarrollo de resultados contextualizados. La experiencia actual mexicana se centra en la rapidez en la emisión de resultados debido a la prioridad administrativa que representa considerar a una evaluación de este tipo como un condicionamiento a una medalla o como elemento necesario para trámites de titulación. Una evaluación del egreso a gran escala comprensiva deberá cuidar entonces los tiempos de su desarrollo, dando prioridad a su adecuada ejecución, más que en diversas aplicaciones al año.

Tradicionalmente los instrumentos se aplican en particular en lápiz y papel, sin embargo, la experiencia nacional e internacional de la evaluación a gran escala está mudando gradualmente a la aplicación por computadora. Esto requiere una fuerte inversión de recursos de todo tipo en el establecimiento de una plataforma acorde con las características de diseño de la prueba, así como la consideración de aspectos operativos, como las particularidades de las computadoras que alberguen el sistema y su disponibilidad dentro de las IES. La posibilidad de llevar a cabo un proceso de este tipo de forma informática favorece posteriormente la conformación de base de datos, la reducción de costos de transportación y almacenamiento de materiales, además que beneficia al uso de reactivos más complejos que los que pueden desarrollarse en lápiz y papel. Un ejemplo del potencial que se puede desarrollar con este sistema de presentación de instrumentos (en particular para la prueba) es el esfuerzo desarrollado por la OCDE en PISA 2015 (OCDE 2015b). El proyecto hace uso de simulaciones mediante el uso de la computadora para generar datos que permitan al estudiante interactuar con el contenido de los reactivos y así potenciar la diversidad y complejidad de respuestas que se pueden obtener.

De igual forma, los cuestionarios pueden administrarse de forma electrónica e, incluso, en línea. Si bien el diseño de los instrumentos debe realizarse con cuidado con el tiempo necesario, los tiempos de ejecución pueden reducirse significativamente mediante el uso de herramientas tecnológicas que actualmente comienzan ya a desarrollarse en proyectos a nivel mundial.

El uso de elementos tecnológicos, como el uso de plataformas digitales o en línea para el desarrollo de la aplicación de los instrumentos favorece también la capacitación de los aplicadores, ya que su función se ve simplificada. De la misma manera, el uso de formatos de control para la aplicación se reduce considerablemente, ya que los sistemas registran asistencia de sustentantes, tiempos de ejecución, así como cualquier otro elemento de interés por parte de la instancia evaluadora. Por supuesto, los tiempos, costos y controles en el transporte de materiales

pueden reducirse de forma importante, dada la factibilidad de reducir al mínimo el uso de papel; por lo que los medios digitales hacen más fluidos estos procesos.

El plan de organización y logística de la aplicación se deberá desarrollar en función de la forma de administración de los instrumentos, así como considerando el presupuesto y tiempos que se pretenda destinar a ello. En la práctica cotidiana de la evaluación las fechas de aplicación determinan la planeación de las actividades de diseño, calificación y resultados. Si bien esto es difícil de modificar, es deseable llevar a cabo una adecuada planeación para que el proyecto no se vea modificado en función de aspectos administrativos, como normalmente suele suceder en la práctica cotidiana.

## **Análisis de la información**

El proceso de análisis de la información en evaluación tiene dos objetivos precisos: el primero de ellos consiste en la identificación de los parámetros psicométricos de los reactivos utilizados, y, el segundo, en identificar propiamente los resultados obtenidos. Ambos son procesos necesarios para la conformación de la información que permitirá retroalimentar a los egresados y a las instituciones. A continuación se presenta una descripción respecto a cada uno de estos objetivos.

### *Análisis psicométrico*

En el caso del primer objetivo, como se comentó en el capítulo previo, se pretende identificar a aquellos reactivos con deficiencias psicométricas respecto a ciertos estándares de medición (como la dificultad, la discriminación y el ajuste al modelo teórico) con el fin de excluirlos de análisis más finos y así mejorar la medición del instrumento utilizado, llevando a cabo de esa forma el proceso de *calibración*. Bajo una perspectiva comprensiva este análisis es también relevante, sin embargo, no con fines únicos de calibración, sino de identificar los posibles errores conceptuales del reactivo o de sesgo que orienten respuestas no esperadas por parte de los representantes disciplinares. Tradicionalmente se considera que el proceso de calibración permite que el instrumento (y, por ende, la evaluación en sí) tenga validez. Prueba de ello es la publicación del INEE (2015), en la cual se hace referencia de la validez como un término propio de los instrumentos (en concreto, de las pruebas y no en particular de los cuestionarios). Paradójicamente en esa misma publicación se menciona que la validez refiere a la forma de difusión de resultados, el uso de los mismos y sus consecuencias. A pesar de esta incongruencia, la validez de la evaluación no debe basarse

únicamente en los parámetros psicométricos de las pruebas utilizadas, sino en cuanto a qué tan válidos son los resultados que se generan para el objetivo de la evaluación y la población que hace uso de ellos.

De esta forma, la validez no se sustenta en los parámetros psicométricos de los reactivos, sino desde el hecho que la evaluación se sustenta en la opinión de los involucrados, considerando una diversidad de aspectos que envuelven al objeto de la evaluación. Adicionalmente, vale hacer uso de los psicométricos que se generen con fines de hacer ajustes a los reactivos utilizados cuando éstos reflejen que no funcionan de forma adecuada en la población de interés. Por supuesto, al igual que en cualquier medición o evaluación de rasgos humanos, no se puede considerar a los parámetros psicométricos de los reactivos como elementos inamovibles, ya que éstos dependen de las respuestas de los estudiantes y las respuestas de los estudiantes dependen de un momento y contexto en particular. Bajo ese supuesto de dinamismo los psicométricos no deberán ser elemento de decisión sobre qué reactivos se deben incluir o no, sino que deben ser usados como un referente para el ajuste de los instrumentos posterior a una aplicación piloto y con fines de mejora para aplicaciones posteriores. Por supuesto, es recomendable llevar un seguimiento de los reactivos conforme se analizan al pasar de las aplicaciones para conocer si los patrones de respuesta se deben al contenido del reactivo, o bien, a condiciones particulares de la población a la que fue aplicada.

Lo relevante del análisis psicométrico bajo la metodología propuesta radica no en la consideración de los aspectos matemáticos como definitorios de la prueba misma (algo similar a como sucede con un puntaje en una prueba, la cual no puede ser definitiva de la persona en sí misma), sino en que los datos pueden entonces considerarse como un indicador de la forma en que los egresados a quienes se les aplicó la prueba reciben la prueba y responden a ella. Por ende, el análisis de estos datos deberá considerar los elementos cuantitativos como insumo para un análisis cualitativo que definirá si el reactivo es pertinente, si tiene algún sesgo conceptual o poblacional o si requiere algún ajuste de redacción o de información que facilite su incorporación a las siguientes versiones del instrumento.

El tipo de análisis psicométrico por utilizar estará mucho en función del tipo de reactivos por analizar, ya sean dicotómicos o politómicos (éstos en particular son apropiados para los cuestionarios, dentro de los cuales cada reactivo suele construirse bajo escalas ordinales), o bien,

si son de respuesta abierta. En este último tipo de reactivos se requiere además un proceso de codificación.

Los proyectos PISA y LLECE, los cuales son un referente directo de experiencia internacional en la cual participa nuestro país, hacen uso de rúbricas holísticas para poder codificar o asignar un valor numérico a las respuestas de los egresados, para con ello posteriormente conformar la base de datos que favorecerá el análisis psicométrico y de resultados de los estudiantes. La experiencia nacional en pruebas utilizadas por Ceneval distintas al EGEL suelen hacer uso de rúbricas analíticas, que si bien dan información detallada, implica un desarrollo más elaborado, pero sobretodo, enlistar una cantidad de criterios tan detallados que puede complicar que los codificadores ejerzan un juicio adecuado.

El desarrollo de las rúbricas para las respuestas abiertas debe considerar a la instancia evaluadora con fines de orientación metodológica, así como representantes disciplinares con experiencia docente, dado que ello permite identificar la gama de posibilidad de respuestas por parte de los estudiantes. Por supuesto, debe considerarse una muestra de respuestas de estudiantes que favorezca el desarrollo de categorías. A continuación se muestra un ejemplo de rúbrica para el reactivo mostrado en la Tabla 5.3.

**Tabla 5.5 Ejemplo de rúbrica para un reactivo de respuesta abierta.**

**Reactivo:**

A mediados de año, un hotel *boutique* de la zona centro de Guadalajara, Jalisco se ve en la necesidad de pronosticar la demanda total de sus servicios de alojamiento, ello con la finalidad de planear su presupuesto para el siguiente año. En años previos han observado que la demanda total se ve influida fuertemente por los *puentes* y fines de semana largos, así como por el festival anual gastronómico que se dan en la región, dado el interés del Gobierno Municipal de fomentar el turismo gastronómico. En particular, estos elementos influyen en un incremento de demanda total de hasta 70% respecto a la temporada baja.

Considerando el objetivo y las variables descritas, ¿cuál de los siguientes métodos permitirá hacer la estimación más precisa de la demanda total del alojamiento?

- A) Análisis de la composición del mercado
- B) Cálculo del índice de factores del mercado

Argumente su respuesta:

---

---

**Rúbrica:**

Elemento por evaluar (objetivo del reactivo): Identificar si el egresado decide la o las técnicas para pronosticar la demanda total de un producto o servicio turístico, considerando una argumentación acorde con el contexto del producto o servicio.

*Códigos posibles*

Código 12. Selecciona “Análisis de la composición del mercado” y da una argumentación que refleje el entendimiento sobre cómo la composición del mercado puede determinar la posible demanda en años posteriores.

Código 11. Selecciona “Cálculo del índice de factores del mercado” y da una argumentación que refleje el entendimiento sobre cómo el desarrollo de un índice que considera los elementos medibles del mercado circundante permitirá estimar la demanda en próximos años.

Código 01. Selecciona “Análisis de la composición del mercado” o “Cálculo del índice de factores del mercado” y da una argumentación que no refleja entendimiento del contexto de mercado del hotel o no refleja un entendimiento del porqué de su selección.

Código 02. Selecciona “Análisis de la composición del mercado” o “Cálculo del índice de factores del mercado” y da una respuesta que no refleja un esfuerzo real para dar respuesta al problema planteado.

*Notas a los códigos*

La selección de la respuesta puede ser omitida, siempre y cuando pueda inferirse dentro de la argumentación del estudiante, por lo que no debe ser una condición necesaria para la asignación de un código en particular.

Fuente: Elaboración propia.

La rúbrica mostrada considera para este ejemplo en particular un código de dos dígitos. El primer dígito de cada código determina el crédito: “0” para respuestas que no merecen crédito y “1” para

respuestas merecedoras de crédito. El segundo dígito de cada código representa una clasificación del tipo de respuesta. Por ejemplo, para las respuestas que merecen crédito, el “2” permite identificar que la respuesta se orientó hacia el análisis de la composición del mercado, por lo que la respuesta del estudiante deberá ser valorada bajo esa selección; lo mismo sucede con el código “1”, cuyas respuestas deben valorarse bajo la condicionante de selección de “Cálculo del índice de factores del mercado”. Ahora bien, en el caso de las respuestas que no merecen crédito, el “1” permite identificar que las respuestas representan un claro esfuerzo de contestar con una argumentación plausible, pero que son insuficientes para considerarse meritorias de crédito; por su parte, el dígito “2” permite distinguir respuestas que no representan un esfuerzo serio por resolver el problema. Esta distinción en las respuestas que no merecen crédito es necesaria, ya que un simple código “0” no permitiría discriminar de las respuestas que representan un esfuerzo legítimo de dar solución pero que no resultan lo suficientemente sólidas para otorgarle crédito, de aquellas respuestas que tienen un interés de burla o algo similar.

En caso de reactivos que incluyan decisiones éticas en su planteamiento, la rúbrica deberá elaborarse con tal cuidado que impida la codificación orientada al *deber ser*, apoyándose fuertemente en los representantes disciplinares, quienes deben conocer a fondo las particularidades éticas que subyacen algunas de las decisiones profesionales a las que se enfrentan cotidianamente. Por supuesto, en estos casos, resulta valioso no solo la identificación del “hacer lo correcto” por parte de los egresados, lo cual sería más un juicio moral, sino valorar la justificación de las decisiones, de la misma forma en que se plantea la rúbrica en el ejemplo previo. Esta forma de plantear la rúbrica permitirá valorar la manera en que los egresados se responsabilizan por las opciones seleccionadas, más que valorar si pueden seleccionar lo que *debe de ser*.

Por supuesto, el análisis de calibración de reactivos de respuesta construida deberá desarrollarse con la misma finalidad mencionada previamente, la de la búsqueda de retroalimentación de los reactivos respecto a su constitución y cómo es que éstos generan determinadas respuestas. El tipo de modelo psicométrico que puede ayudar al desarrollo de este tipo de análisis es la TRI orientado a reactivos politómicos (Hontagas, Ponsoda, Olea y Revuelta, 1998).

Evidentemente el desarrollo de las rúbricas no es un proceso sencillo, por lo que requiere forzosamente de la participación de diversos agentes. Es necesaria una orientación metodológica por parte de la instancia evaluadora, así como la aportación conceptual que solo los profesionales

y docentes de la disciplina pueden aportar. Por supuesto, también implica la consideración de erogación de recursos y un tiempo necesario para que éste sea un trabajo adecuado. El proceso mismo de desarrollo de rúbricas requiere una planeación estratégica como se piensa normalmente la redacción de reactivos para la prueba y los cuestionarios, ello bajo una orientación metodológica constante.

De la misma forma, la codificación de respuestas debe ser desarrollada necesariamente por representantes disciplinares, es decir, profesionistas y/o docentes contratados *ex profeso* para dicha labor. Este procedimiento también puede verse beneficiado debido a aplicaciones computarizadas, ya que facilita la manipulación de información y de su codificación directa de forma digital, lo cual, por supuesto, agilizará la conformación de la base de datos al no considerar un proceso de captura de información de cuadernillos en papel. Incluso, dado que se hace uso del juicio humano para la emisión de los códigos a cada una de las respuestas de los egresados, se sugiere la consideración de estrategias de verificación de consistencia de codificación. Ello puede hacerse mediante la doble revisión (o en ocasiones triple) de algunas respuestas, con fines del cálculo de la consistencia de juicios, lo cual permite asegurar, al menos de forma relativa, que los codificadores están emitiendo su juicio en función de los criterios establecidos en las rúbricas de cada uno de los reactivos de respuesta abierta. Por supuesto, la consistencia no dependerá tanto de los jueces, sino de las reglas de codificación establecidas en las rúbricas mismas, las cuales deben ser lo más clara posibles para evitar confusiones e inconsistencias entre las valoraciones que se realicen sobre las respuestas analizadas; por supuesto, una forma de asegurar esta consistencia también será mediante la capacitación que se realice como forma de entrenamiento de los jueces participantes.

#### *Análisis de resultados*

Una vez concluido el proceso de análisis psicométrico de reactivos y habiendo realizado ajustes a los instrumentos para un mejor entendimiento por parte de los egresados, docentes y directivos, es necesario el desarrollo del análisis de datos para efecto de resultados. Este resulta ser uno de los momentos más complejos y cruciales de la información. En las instancias evaluadoras hay áreas completas dedicadas a realizar esta labor, siempre bajo estándares estadísticos robustos que permitan *asegurar* que la información que se emite en reportes de resultados es lo más fidedigna posible.

Para una evaluación de este tipo de forma comprensiva, la emisión de resultados es prioritaria, pero no debido a la importancia de la solidez de los métodos matemáticos que permitan dar una supuesta objetividad de los datos, sino porque debe planearse una emisión de resultados que considere la contextualización de los datos y que sea de gran utilidad para retroalimentar a los participantes y a las instituciones encargadas de administrar los programas educativos en cuestión. Bajo esta lógica, se han desarrollado reportes complejos mediante modelos multinivel (Murillo, 2008) o los modelos de ecuaciones estructurales (Ruiz, Pardo y San Martín, 2010). Baste con revisar reportes de evaluaciones emitidos por el INEE tales como *Análisis multinivel de la calidad educativa en México ante los datos de PISA 2006* (INEE, 2008), *Análisis de los resultados de la prueba nacional de aprovechamiento en lectura en secundaria: estudio multinivel de logro y tendencias* (INEE, s/f), *Perfil de los lectores de los estudiantes de secundaria de PISA* (Tristán, 2012) para valorar la forma en que este tipo de análisis permiten construir modelos de interpretación considerando una diversidad de variables incluidas en los cuestionarios. Este tipo de análisis, por supuesto, tienen una fuerza importante en cuanto se busca explicar ciertos resultados educativos, los cuales deben rescatarse, ya que favorecen la comprensión de lo observado. Sin embargo, un paso previo consiste en la determinación de puntos de corte, ya que es ese proceso que permitirá un primer entendimiento de lo observado y su posterior relación con el contexto.

Como se analizó en el capítulo previo, la experiencia nacional del EGEL promovió el desarrollo de una metodología criterial con fines de mantener un estándar relativamente *estable* que diera a las IES una perspectiva de objetividad y sistematización que facilitara la comparación de resultados, de tal suerte que si se otorgan medallas o alicientes a los estudiantes, entonces el método de comparación debe ser el mismo para todos, independientemente de sus condiciones. En la actualidad es rara vez aplicada la evaluación normativa, dada una tendencia del uso de mediciones criterios con fines de incentivar el trato *objetivo* de las personas evaluadas, controlando así los sesgos que pueda tener una posible apreciación subjetiva de los datos (Álvarez, 2001). Esta perspectiva de medida puede representar diversos inconvenientes pensado bajo una perspectiva comprensiva. Al respecto, Álvarez (2001, pág. 18) comenta lo siguiente:

“Operativamente, la estructura de las pruebas objetivas consiste en un enunciado que especifica en términos precisos de los logros previstos. Debe estar redactado de modo tan exacto que solo admite una única interpretación que da por válida una única respuesta. De esta forma y sin explicitar los valores formativos que se sacrifican en el intento –los procesos complejos de elaboración principalmente–, ni la ideología conservadora que estas pruebas representan –nacieron con claros propósitos de control y de selección–, evita

técnicamente interpretaciones que dan pie a respuestas que no estén controladas con anterioridad.

El conocimiento factual o los niveles de cognición elementales constituyen el objeto de estas pruebas. Más difícil, sino imposible, lo tienen cuando el objeto de conocimiento es el pensamiento crítico o creativo, o cuando se trata de hallar o aplicar solución a problemas, o dar respuestas en situaciones no previstas. Difícil también cuando se trata de evaluar aprendizaje significativo, a no ser que la significación venga marcada por criterios ajenos al sujeto que aprende, lo que representa una contradicción en sus términos. Pero estas consideraciones no entran en ese proceder. La preocupación se centra en el control técnico de los resultados y del rigor en la aplicación de la técnica, que después de sacrificar los valores resistentes al control técnico, garantiza el trato objetivo en la medición de los resultados”.

Este problema se analizó en el capítulo previo, en éste se indicaba que la metodología del EGEL para el establecimiento de criterios orillaba a considerar en demasía el impacto de la perspectiva del Consejo Técnico y la proporción de estudiantes en cada nivel de desempeño (convirtiéndola así en una decisión con una fuerte carga normativa). Adicionalmente vale considerar las argumentaciones de Álvarez (2001) respecto a la pérdida de sentido pedagógico y de retroalimentación de la medición criterial, la cual sobrepone los elementos técnicos a las características propias de los egresados.

Procurando una consistencia epistemológica y metodológica, en la presente propuesta se sugiere el uso de una calificación normativa, la cual es coincidente con la pedagogía de la complejidad, ya que no solo los elementos (en este caso, la respuesta de los estudiantes) están conectados y dependen el uno del otro (Morin, 2004), sino que también las variables contextuales entran en esa dinámica; además coincide con la perspectiva comprensiva en el entendido que la evaluación no debe representar un pensamiento lineal ni completamente estable. Cada aplicación y, por ende, los resultados que se emiten, son únicos y no pueden ser generalizados. Las acciones humanas deben ser contextualizadas y cada respuesta a los instrumentos no se da únicamente solo por el aprendizaje logrado, sino dada una variedad de factores, entre ellas, el grupo de personas que presentan la prueba en un mismo momento, ya que coinciden en un momento histórico preciso. Considerar una medición criterial implicaría asumir que hay *un* estándar en particular que todos los estudiantes deben alcanzar, sin importar necesariamente sus planes de estudio o condiciones bajo las cuales desarrollaron sus estudios. Por supuesto, una evaluación requiere necesariamente de un estándar, ya que su uso es inherente a la evaluación (Stake, 2006), ya que es a través de

ellos que se establece un referente respecto a lo que se tiene que valorar, no son más que el proceso mediante el cual se establece el encuadre de una retroalimentación. Sin embargo, la medición criterial implica una insistencia respecto al establecimiento de un estándar fijo, inamovible, con fines de asegurar una relativa objetividad. La insistencia en el uso de la medición criterial es, por supuesto, una consecuencia del interés de mantener una valoración del ser humano como un producto y no como una experiencia. Por supuesto, el valor pedagógico de esta visión es seriamente limitado. Adicionalmente no debe olvidarse que todos los egresados y sus condiciones no son iguales, por lo que deberán valorarse de forma independiente; ello favorece lo que el INEE (2014), bajo un enfoque de derechos, establece como equitativo: la trascendencia de la igualdad, aplicando criterios diferenciados. En ese sentido, la emisión de resultados mediante un método normativo podría favorecer a conocer con mayor profundidad las causas que originan los resultados, para que con ello las IES puedan establecer las acciones de mejora que consideren pertinentes.

Por supuesto, la búsqueda de comparación entre instituciones se vuelve necesaria en un contexto centrado en la rendición de cuentas. Si bien el uso de una medición criterial ofrece una artificial pero satisfactoria comparación entre instituciones, no es lo más recomendable para fines pedagógicos (aunque sí para fines de competencia). Como alternativa siempre es necesario especificar el contexto de los resultados. Si únicamente se establece un ranqueo de resultados, los datos, sin contexto, serán comparables, pero, en la realidad contextualizada, los datos de un ranqueo carecen de un sentido educativo. Como ejemplo se puede pensar en la obtención de un título de licenciatura, para la SEP, la emisión de un título será válida siempre y cuando se cumplan con los requisitos de un Plan de estudios con ROVE (y para efectos legales y laborales lo es), pero contextualizando los datos, no se pueden valorar las condiciones de la obtención de un título para un estudiante indígena de una zona marginada en comparación con el de alguien que tenga acceso a universidad privada; no solo por las condiciones de acceso, sino por los materiales con los que realizan los estudios e incluso las condiciones de titulación (hay escuelas que no requieren de trabajos recepcionales, como las tesis). Lo mismo sucederá a gran escala, las condiciones de cada contexto tendrá cierta influencia en las condiciones de sus egresados. Por ende, una medición normativa parecerá más apropiada, ya que favorece la sensibilidad al momento de aplicación. Esta metodología va acorde con la epistemología de la complejidad, ya que esta perspectiva implica considerar lo incierto dentro del conocimiento (Brower, 2009), lo cual permite un retorno constante entre certeza e incertidumbre. La complejidad de un fenómeno educativo, como en este

caso, el aprendizaje de los estudiantes, radica principalmente en su dinamismo y que siempre se encuentra en expansión; ese dinamismo puede ser retratado de forma más apropiada mediante la metodología normativa. Cabe aclarar que no se trata de generar una propuesta que flexibilice criterios para ciertas zonas o contextos, sino de emitir estándares de comparación considerando las condiciones bajo las cuales se da la formación, lo cual favorecerá el entendimiento del porqué de los resultados, facilitando la implementación de políticas y acciones dentro de las IES para la mejora de la formación de los estudiantes.

Bajo esta mirada, el establecimiento de puntos de corte no deberá establecer metodologías criterioales, como las descritas en el capítulo previo, sino que deberá valorar en mayor medida los resultados de los egresados en función de diversas variables de contexto en particular y, por supuesto, hacer uso de la perspectiva experta de los representantes disciplinares. La medición normativa se enfoca principalmente al uso de la distribución de frecuencias para el establecimiento de los puntos de corte, dando especial prioridad a la media de puntajes y las unidades de desviación estándar para determinar los diferentes niveles de desempeño. Pareciera un método anticuado y rústico, dada la aparente robustez de los métodos criterioales, pero no lo es; bajo la perspectiva epistemológica que ha servido de sostén de esta propuesta resulta una metodología válida y coherente con los principios de diversidad, de atender al contexto y de considerar los resultados de una evaluación como solo una mirada fotográfica, la cual no puede ser determinante del ser humano, dado su propia existencia dinámica. La definición de los niveles de desempeño sobre la prueba requerirá la agrupación de los estudiantes de acuerdo con las características contextuales que sean establecidas por los representantes disciplinares (en el entendido que su importancia se entiende dentro de la disciplina misma); ello implicará, por tanto, que los niveles de desempeño consideren el contexto de los egresados, por lo que la comparabilidad de datos a los que la evaluación a gran escala tiene acostumbrado puede verse seriamente amenazado.

Por supuesto, estas condiciones tienen un efecto directo sobre el tipo de análisis para la emisión de resultados. Se requiere entonces del uso de estadísticos multinivel para desarrollar modelos que permiten resumir las características de los egresados en función de las variables contextuales medidas en los cuestionarios. Para ello se requiere el cumplimiento de los supuestos de la estadística inferencial, aunque ello no necesariamente podrá ser posible, dado que la población no

necesariamente es seleccionada al azar y su objetivo no es generalizar a toda la población, sino la caracterización de los egresados que participen en cada aplicación anual.

La consideración de análisis complejos cuyos resultados se sustentan en dar un papel relevante a las variables de contexto que definen a una generación de egresados en particular permitirá entonces conjuntar una diversidad de perspectivas y elementos de decisión de los egresados frente a las situaciones que la prueba les presenta y las condiciones que tuvieron que enfrentar (reconocida en los cuestionarios y retomada por los análisis multinivel). Si bien la presente propuesta no presenta a detalle la caracterización metodológica de este tipo de análisis estadístico, resulta una forma viable de conjuntar de forma cuantitativa la diversidad que esta propuesta comprensiva ha resaltado. Si bien, como se comentó al inicio del capítulo, hay opciones que tienden a la triangulación, la presente propuesta, enfocada en elementos a gran escala, debe buscar métodos más parsimoniosos para el tipo de datos que se recolectan. En particular, los datos de juicio en la prueba y de opinión en los cuestionarios, los cuales se codifican con fines de tratamiento estadístico y de su inclusión en los resultados. Por ende, el último reto de esta propuesta consiste en la comunicación de esos resultados.

## Emisión de resultados

Cualquier proceso de evaluación carecería de sentido sin considerar el proceso de emisión de resultados, ya que no habría oportunidad de decir nada respecto a lo que se analizó y los recursos, tiempo y esfuerzo invertido serían infructuosos. Esta preocupación se acrecienta en una evaluación bajo el enfoque comprensivo, dado que la comunicación con los *stakeholders* y los evaluados mismos es primordial dado que pensar a la evaluación como una *retroalimentación* hace de este paso uno de los más cruciales. En la emisión de los resultados es donde se reafirma la perspectiva de la evaluación, por lo que su elaboración debe cuidar los aspectos propios de la evaluación comprensiva que se han discutido a lo largo del presente trabajo.

Como se analizó en el Capítulo 2, hay toda una diversidad de evaluaciones a gran escala y cada una de ellas establece una forma de emitir sus resultados, algunos serán públicos y otros serán privados, pero los resultados siempre estarán en función del objetivo de la evaluación. En este caso, el objetivo está enfocado a la retroalimentación de los programas de estudios que participen en el proyecto, así como en que los egresados participantes, quienes se espera que conozcan el estatus de su preparación en función de los límites conceptuales y de la praxis de su disciplina

enmarcada en la evaluación. Siendo así, la retroalimentación debe ir en dos sentidos: el primero de ellos dirigido a las IES, responsables de los planes de estudios participantes y, el segundo, dirigido a los egresados participantes. A continuación se describen algunas propuestas en coincidencia con los objetivos de la evaluación.

#### *Retroalimentación a las IES*

La emisión de resultados dirigidos a las IES es una experiencia que ya se ha dado en México gracias a los EGEL, por lo que las instituciones están acostumbradas a recibirlos, y no solo eso, sino a hacer uso de ellos para efectos de su uso en la generación de indicadores de eficiencia terminal y titulación para los fines que a cada programa educativo convienen. De acuerdo con lo comentado en el capítulo previo, los informes de resultados a las IES consideran un conjunto de estadísticos descriptivos, en particular, solo la proporción de estudiantes conforme a un nivel de desempeño preestablecido. Sin embargo, bajo una perspectiva comprensiva la forma de comunicación de los resultados deberá considerar más que solo un conjunto de estadísticos descriptivos.

En la presentación de resultados de evaluaciones, las herramientas auxiliares de análisis, como las tablas y las gráficas, no deberán ser lo prioritario en los resultados, aunque, por supuesto, deberán incluirse. Los informes deberán ser lo más ejecutivos posibles y deberán describir de forma breve las condicionantes contextuales de mayor relevancia y aclarar al lector de los límites de la evaluación. La primera limitante es la consideración que la evaluación anual será una especie de *foto instantánea* de los egresados participantes. Ello implica evitar considerar a la evaluación como un grabado en piedra que determine lo que una persona será capaz de hacer o no hacer en su ejercicio profesional en el resto de su vida. La evaluación servirá entonces para conocer las características de los egresados en el instante (idealmente ese instante será al momento de concluir sus estudios profesionales) y poder considerar los resultados en los ajustes a los aspectos institucionales que favorezcan a próximas generaciones.

Otra limitante es la consideración de variables de contexto para delimitar resultados. La evaluación a gran escala ha comunicado sus resultados de una manera simple con fines de entendimiento de las personas que regularmente están ajenas a procesos metodológicos para su elaboración, sin embargo, no se ha logrado transmitir que los resultados siempre estarán en función de todo lo que sucede alrededor de los estudiantes; por tanto, la comparabilidad de los resultados de un estudiante con cualquier otro estudiante del país se pone seriamente en duda.

Corresponde a la instancia evaluadora la descripción de las limitantes de sus juicios en el momento de emisión de informes de resultados.

Adicional a la descripción de limitantes, la comunicación de resultados deberá describir las características observadas de los egresados que participaron en esa aplicación, por supuesto, previa descripción breve<sup>27</sup> de la metodológica utilizada. La descripción de resultados puede organizarse mediante las áreas valoradas, procurando mantener un equilibrio entre áreas de oportunidad (por supuesto, con una adecuada descripción del contexto) y las potencialidades que pudieran haber generado el conjunto de egresados participantes. Esta descripción puede desarrollarse con fines de una lectura fluida mediante el uso de infografías y gráficas.

Un documento que privilegie la comunicación fluida no necesariamente implica dejar a un lado el rigor metodológico, el cual puede explicarse de forma detallada en informes técnicos independientes, cuya audiencia se centre en evaluadores profesionales e investigadores. Sin embargo, continuando en la línea comprensiva, los resultados de la evaluación deben estar centrados en los *stakeholders* y sus necesidades. De esta forma, los informes institucionales deben ser fluidos, entendibles y comunicar de forma efectiva y equilibrada las características de los egresados. Por supuesto, dada la necesidad institucional de contar con los estadísticos descriptivos de los egresados con fines administrativos, éstos pueden facilitarse para cada institución. Lo relevante es mostrar a las IES la posibilidad de contar con la información necesaria que no solo les permita continuar con sus gestiones y necesidades particulares, sino también una retroalimentación auténtica para efectos de la mejora de la práctica educativa en beneficio de las generaciones por venir.

Desarrollar este tipo de reportes no es sencillo ni rápido, dado que debe considerar previamente el análisis desarrollado y no bastará únicamente con establecer cantidades y proporciones de egresados por nivel de desempeño, sino la redacción de los resultados y su manejo en infografías mediante profesionales de la educación y del diseño. A pesar de ello, la evaluación debe ser de utilidad y, dado que la evaluación es una *instantánea*, el tiempo de emisión de resultados será el recurso más valioso de todo el proceso de evaluación. Si la instancia evaluadora falla en su determinación, la evaluación bajo esta lógica comprensiva perderá mucho de su sentido.

---

<sup>27</sup> Dado que este tipo de informes deben ser ejecutivos con fines de no distraer al lector de su verdadero interés (los resultados), se sugiere que la descripción metodológica sea concisa y breve.

Para afrontar este problema es necesario evitar dar un informe por institución. Si bien como anexo se pueden considerar datos cuantitativos como las tradicionales proporciones para cada IES, se debe emitir un informe (con todos los elementos descritos) para el conjunto de IES que se hayan tomado en cuenta para el establecimiento de puntos de corte y el análisis de resultados, es decir, un reporte que englobe a los egresados de las instituciones que en su conjunto definieron los resultados, dado ello a través de su similitud contextual conocida mediante los cuestionarios; ello facilitará la comparación de las IES con otras que compartan elementos de contexto y considerando de esa manera elementos que favorezcan el desarrollo de sus egresados.

El mayor reto de la comunicación de resultados a las IES resulta en otorgar un insumo relevante que permita que las autoridades educativas obtengan información que les resulte pertinente para poder implementar políticas y acciones concretas que favorezcan la formación de los estudiantes. Para ello se hace necesario no solo que conozcan la proporción de estudiantes en los niveles de desempeño definidos, sino las condiciones contextuales (dentro de las cuales se incluyen elementos de infraestructura institucional, o bien, del manejo de contenidos por parte de los docentes, entre otros factores que puedan ser indagados mediante los cuestionarios) que determinan cómo es que los egresados se están formando en elementos disciplinares y éticos. Por lo tanto, los informes de resultados deberán mostrar una clara línea de trabajo, evitando centrarse solo en los resultados de aprendizaje, sino explicando las condiciones de formación que experimentaron los egresados.

#### *Retroalimentación a los egresados*

Como segundo objetivo, los egresados están en derecho de conocer sus resultados. Por supuesto, una vez que se han dado resultados por IES se puede proceder al desarrollo de informes individuales; ello no por el hecho que las IES sean más relevantes, sino porque para emitir resultados por persona se deben tener establecidas previamente las variables que permiten agrupar estudiantes e instituciones, de otra manera los resultados serían descontextualizados.

El informe de resultados de la experiencia del EGEL presenta el dato numérico y su equivalente en una *etiqueta* de desempeño (aún no satisfactorio, satisfactorio o sobresaliente). En el caso de la presente propuesta los resultados individuales deberán contener mediante una muy breve descripción las áreas evaluadas, resaltando sus potencialidades y áreas de mejora. Por supuesto, el uso de gráficas sencillas de interpretar también se recomienda. Incluso, si la infraestructura lo

permite, se puede contar con un repositorio de resultados en línea, en el cual cada estudiante pueda revisar sus resultados mediante el ingreso de su número de registro. Muchas instituciones, tales como la UNAM o el Ceneval, cuentan ya con registros de este tipo en sus respectivas páginas de Internet.

En particular no se sugiere el uso de “testimonios” de desempeño alguno, ya que éstos no cumplen una función de retroalimentación y no ayudan en sí a la búsqueda de la mejora de las diversas áreas evaluadas, además que antepone un mérito o recompensa en la evaluación; rompiendo así con el sentido pedagógico de la evaluación comprensiva. La intención es mostrar en los reportes una descripción breve, ya sea en forma de prosa, o bien, mediante el uso de puntajes<sup>28</sup>, buscando una lectura apropiada a los ojos de los estudiantes.

Los reportes para los estudiantes deberán ser breves, entendibles y lo menos metodológicos y técnicos posible, considerando que el interés de comunicación en este caso es el de resaltar elementos que favorecerán el desarrollo profesional del egresado, así como elementos que puede mejorar con el mismo fin. Por supuesto, el juicio sobre los estudiantes estará supeditado a los resultados dirigidos a las IES, ya que consideran un contexto y un grupo de instituciones en particular, por lo que solo podrá haber comparabilidad de puntajes solamente entre aquellos estudiantes del grupo en el cual hayan participado en la definición de puntos de corte. Si bien esta descripción metodológica no es necesario que esté explicada dentro de un informe individual, sí es importante que se aclare la imposibilidad de comparación entre grupos de definición de puntos de corte, así como entre aplicaciones. Este punto es relevante, en el sentido que comúnmente se entiende que los resultados de una evaluación son en sí mismos definitorios de una persona (o de un grupo de personas), sin embargo, es menester comunicar a las instituciones y a las personas que los resultados de la evaluación son válidos para el momento en particular. Si el año siguiente un egresado vuelve a presentar la evaluación, sus resultados deben ser diferentes, dada la experiencia profesional y de vida que puede experimentar. Considerando las perspectivas epistemológica y metodológica utilizadas para sustentar el presente trabajo, la comunicación de la incertidumbre de los resultados para con el futuro y entre grupos de estudiantes es necesaria y sana, ya que favorecerá la interpretación de la evaluación como una *instantánea* de los estudiantes que egresen en ese momento.

---

<sup>28</sup> El uso de puntajes no necesariamente significa que no se tengan elementos cualitativos para su interpretación, pero éstos se incorporan dentro de la descripción genérica de las escalas utilizadas.

En ambos objetivos (informes a las IES y a los egresados) se puede considerar el uso de informes electrónicos, dada su facilidad de emisión, distribución y la gama de posibilidades visuales y de conexión entre elementos que ofrecen los formatos digitales. Ello facilitará también la fluidez de la comunicación y el contacto más estrecho con los diversos *stakeholders*, así como abaratar procesos de comunicación. De esta forma, la comunicación debe atender la necesidad de las instituciones respecto al manejo de sus indicadores y recursos frente al ámbito actual de bajo presupuesto público para la educación superior, para lo cual los resultados de evaluaciones de este tipo tienen gran impacto, como ya se ha analizado; asimismo, se privilegia la retroalimentación respecto al estatus de los egresados para un conjunto de instituciones que compartan determinados elementos contextuales, lo cual favorecerá la comprensión del porqué de los resultados que se obtengan y permitirá que los programas educativos tengan como referente un contexto similar al que ellos presentan; por otra parte, favorece que los estudiantes obtengan no una medalla o un certificado, sino una retroalimentación respecto a lo que saben realizar de forma adecuada, pero además, de los elementos en los cuales pueden mejorar como profesionistas, ello bajo un marco del *perfil esperado del egresado* que se ha definido de forma comprensiva mediante la participación de una enorme diversidad de *stakeholders*.

Esto, por supuesto, da un sentido de profundidad a los resultados, ya que permiten atender diversos aspectos, resaltando la retroalimentación –pieza fundamental de una evaluación comprensiva– sin descuidar los elementos administrativos que las IES requieren bajo la política basada en la rendición de cuentas en la que actualmente se mantiene el SES.

## Conclusiones

La evaluación del egreso de la licenciatura en México ha sido desde hace dos décadas uno de los elementos estratégicos de implementación de una cultura de evaluación y de rendición de cuentas para –de acuerdo con el discurso oficial– fomentar mejoras que beneficien la formación de los profesionistas de nuestro país, ello como uno de los proyectos complementarios al establecimiento del sistema de calidad de la educación superior. Por supuesto, la política educativa mexicana no ha estado ajena a la proliferación de servicios de evaluación a gran escala que se ha dado en todo el mundo como herramienta que favorece la implementación de una perspectiva basada en la competencia bajo la influencia de un libre mercado. La elaboración de proyectos de evaluación a gran escala se ha desarrollado también gracias a la posibilidad metodológica que la perspectiva tyleriana de la evaluación permitió, centrando sus esfuerzos en la observación de resultados educativos esperados de un plan de estudios, manteniendo así una dinámica estrecha entre: apertura de mercado nacional – fomento de competencia (y su correspondiente fomento en educación superior) – búsqueda de calidad – evaluación – financiamiento.

Esta dinámica ha favorecido la consolidación de la evaluación a gran escala dentro de la vida académica y administrativa de las IES mexicanas, y en particular a la correspondiente al egreso del nivel licenciatura; sin embargo, como en cada acción científica, es necesario cuestionar los procesos de su desarrollo con fines de mejora o de planteamiento de una nueva postura que favorezca los elementos que las propuestas iniciales no necesariamente han considerado, ya sea por la forma en que los trabajos se iniciaron, o bien debido al momento histórico, social, económico y político que favorecieron sus comienzos, el cual, como es normal, sufre modificaciones con el paso del tiempo. La evaluación a gran escala del egreso de la licenciatura ha llegado a un punto tal de desarrollo que sus mejoras se observan notablemente en los aspectos de medición, en particular con lo relacionado a su precisión y validez en términos psicométricos. Se han logrado desarrollar en diversas partes del mundo diversos instrumentos a gran escala que consideran metodologías de muestreo y de análisis estadístico que favorecen la construcción de modelos explicativos de lo observado y, más aún, modelos matemáticos del funcionamiento de la prueba misma, atendiendo de forma psicométrica elementos como el factor adivinación o incluso el fenómeno de la copia entre alumnos. De esta manera la tradición positivista que mantiene las

raíces de un trabajo de tal importancia en la política educativa y de un fuerte impacto en la vida de los estudiantes en todo el mundo, y en particular en nuestro país, ha favorecido el establecimiento de una madurez tal que se ve reflejada en lo robusto de sus métodos y en el nivel de experticia que se requiere de un profesional para participar dentro de la vida evaluativa. Sin embargo, es justo dentro de esta fortaleza metodológica en donde se encuentra su elemento más débil. Es el rigor metodológico enfocado en la perspectiva que la vio nacer lo cual se ha encargado de establecer fuertes y rígidas barreras epistemológicas y metodológicas a su alrededor, lo cual ha limitado su expansión, pero ha aumentado su capacidad de ahondar en la especialización de la medida de rasgos de aprendizaje.

La evaluación a gran escala enfocada en el egreso de la licenciatura ha alcanzado tal madurez metodológica que le ha impedido desarrollarse bajo otra perspectiva, como sí ha sucedido con otras áreas de la evaluación, como en el caso de la evaluación de programas educativos, área en la cual se han ampliado los horizontes de enfoque y de acción metodológica. Si existen una gran diversidad de enfoques evaluativos, ¿qué ha limitado a la evaluación del egreso de la licenciatura a gran escala a desarrollarse de la misma manera?

Evidentemente, la dinámica de la vida financiera de las IES en nuestro país ha sido un elemento clave en la consolidación de este tipo de evaluaciones, la cual ha favorecido una política educativa enfocada en la rendición de cuentas y, por ende, en un abuso respecto al interés en los resultados de logro académico de los estudiantes; claro es que los resultados de los estudiantes permiten cumplir con la política educativa que la economía del país ha impuesto, sin embargo, la propuesta presentada resulta un primer acercamiento a valorar una posibilidad de desarrollar este tipo de evaluación bajo una metodología más robusta, que también representa una opción de retroalimentación para los que participan de este tipo de ejercicio. Por supuesto, basta con analizar la cantidad de propuestas de evaluaciones a gran escala para darse cuenta que los esfuerzos que se hacen en todo el mundo en este rubro de la evaluación se basan mayoritariamente en una sola perspectiva. Este tipo de evaluación ha considerado como prioritaria la estandarización, es decir, la fijación de metodologías de diseño, calificación y emisión de resultados que sean ajenos a aquello que se asemeje a la diversidad. Resulta una paradoja el hecho que una evaluación que se enfoca en la diversidad que representan los fenómenos a gran escala busque mediante su metodología el establecimiento de límites conceptuales que delimiten esa diversidad a un solo canal de apreciación de los egresados. Sin embargo, esa paradoja ha sido

propiamente su elemento más atractivo, ya que favorece a las visiones educativas con una cierta aversión a la diversidad que caracteriza a este tipo de fenómenos, dadas las condiciones en las cuales se desarrolla un egresado de nivel profesional.

De entre las posibles opciones de enfoques de evaluación, la presente propuesta se enfocó en la evaluación comprensiva propuesta por Robert Stake. El enfoque comprensivo ha sido aplicado particularmente en la evaluación de programas educativos, dado su centro específicamente enfocado a la observación naturalista e interés particular en las cuestiones de proceso y que el evaluador valore la experiencia de *estar ahí*. Esta perspectiva si bien representaba un reto operativo importante, es un enfoque mucho más acercado a la realidad de la pedagogía que se basa particularmente en elementos de complejidad. Las dinámicas educativas de la actualidad requieren dejar atrás el modelo tyleriano de enseñanza y mudar poco a poco a enfoques y metodologías con mayor apertura hacia la diversidad y la complejidad, evitando limitar la praxis pedagógica hacia un enfoque que favoreció fuertemente la enseñanza y la evaluación, pero que a este punto solamente representa ventajas administrativas y económicas para las instituciones encargadas de la formación de los futuros profesionistas. El modelo tyleriano de la evaluación a gran escala ha dado de sí, académicamente hablando, y no ha favorecido el desarrollo de métodos que permitan la observación de elementos que la perspectiva misma no admite. Lamentablemente, como se encontró en la revisión bibliográfica, las propuestas de evaluación no necesariamente consideran la explicitación de la perspectiva que fundamenta su diseño metodológico.

Frente a esta demanda actual de la pedagogía por la consideración de métodos más centrados en la diversidad y complejidad, la perspectiva comprensiva de la evaluación representa una alternativa viable. Por supuesto, la planeación de una propuesta de una evaluación que se desarrolló bajo una lógica tyleriana y con fines políticos perfectamente establecidos no sería un aspecto sencillo. En ese sentido, la consideración de la experiencia nacional de este tipo de evaluación resultó de gran valía, ya que sentó las bases de un trabajo de este estilo sin una cultura previa de la evaluación en México. La consideración del EGEL como caso de análisis permitió conocer la forma de trabajo de este tipo de evaluación en nuestro país (por supuesto, bajo la perspectiva tyleriana tradicional), así como identificar los aspectos que pueden rediseñarse bajo una perspectiva comprensiva. Para este análisis fue necesario considerar que el EGEL se inició en su momento como una solución a los problemas que la evaluación a gran escala representaba para

un conjunto de instituciones educativas reacias a una evaluación y, en particular, a una evaluación externa.

En ese sentido, la consideración de alicientes a la evaluación de los estudiantes representó una valiosa herramienta para la implementación de una cultura evaluativa y la consolidación de una política educativa basada en la competencia y la rendición de cuentas; sin embargo, dicho modelo no ha mutado en los 20 años de este proyecto, lo cual hacía necesaria una valoración de su método, no solo debido a la perspectiva clásica de obtener premios o castigos en función de los resultados de la evaluación, sino también en función del ya comentado avance paradigmático de la pedagogía, que actualmente considera la valoración de otros aspectos y, en particular, de la visión de la evaluación como un sinónimo de la retroalimentación.

Esta perspectiva de retroalimentación es la que subyace a la visión comprensiva, además de la consideración de diversos agentes que fundamenten una evaluación que compenetre a los agentes involucrados y no represente una imposición de una realidad enmarcada por el agente evaluador. Por supuesto, no se debe dejar a un lado todos los avances que el proyecto EGEL ha sentado para una evaluación de este tipo, desde la cimentación de una cultura de evaluación dentro de las IES, hasta la definición metodológica de su desarrollo y los complejos procesos de aplicación y calificación de los datos. Esta experiencia resulta una columna importante para cualquier práctica de evaluación a gran escala del egreso del nivel superior que se pretenda generar.

La presente propuesta no pretende, por supuesto, persuadir a las instancias evaluadoras, como el Ceneval, de desarrollarla e incorporarla dentro de sus objetivos, es un ejercicio de plantear la posibilidad de establecer una metodología distinta para desarrollar una labor de evaluación que resulta importante para la vida académica, así como para la definición de políticas y acciones educativas dentro de la vida de las IES de nuestro país. En ese sentido, la presente propuesta de evaluación consideró la experiencia nacional para valorar un panorama actual de cómo y por qué se generan este tipo de programas; y, como se pudo observar, la propuesta ataca principalmente la necesidad de encuadrar una postura epistemológica firme que de sentido a todo el método utilizado. Normalmente cuando se piensa en *evaluación a gran escala* se hace referencia a elementos técnicos y operativos que, si bien son importantes, deben estar fundamentados en perspectivas epistemológicas que den un sentido y un encuadre a los resultados que una persona

recibe sobre su formación. Al observar que las experiencias conocidas al respecto carecen de una fundamentación explícita de su eje, resultaba necesario dar especial atención a esa cuestión.

De esta forma, la propuesta procura mantener una ilación lógica en función del encuadre del enfoque comprensivo de la evaluación. Evidentemente hay elementos que resultan tan operativos que si bien pueden repensarse en función de esta perspectiva, también pueden mantenerse prácticamente sin cambios como cuando se hace mediante la perspectiva tradicional. El ejemplo claro de ello es la consideración de una desagregación de los elementos por evaluar, si bien no se sugiere una atomización del conocimiento tal como se realiza en el EGEL, es necesario plantear un enunciado que permita la medición de los elementos de conocimientos, habilidades y decisiones éticas que interese evaluar. Esta desagregación de las áreas de interés es característica de una metodología completamente acorde con la perspectiva psicométrica y la visión tyleriana de la educación; a pesar de ello, como se comentó, la medición de los rasgos psicológicos de interés (por ejemplo, opinión y aprendizaje) es necesario dentro de un proceso de evaluación, por lo que la diferencia en esta propuesta radica principalmente en el protagonismo que la medición tiene para el proyecto de evaluación, así como en el nivel de atomización con el que se aborda el contenido por evaluar. En este caso, la desagregación solo se realiza con fines de facilitar el desarrollo de instrumentos, pero no representa en lo absoluto el insumo principal de validez que la perspectiva psicométrica defiende.

En otro aspecto, el presente trabajo resalta el hecho de hacer uso de los cuestionarios como instrumentos igual de valiosos que las pruebas. Si bien los cuestionarios no pretenden *probar*, sino contextualizar, son formas de acercamiento a la realidad de los egresados que permiten poner en situación lo que se pueda analizar mediante las pruebas, así como dar un sentido a la retroalimentación que surja de la evaluación, como se comentó en el último capítulo. Bajo esta lógica son los cuestionarios los que dan sentido a la evaluación y permiten reafirmar la validez de lo que la instancia evaluadora pueda decir respecto de los egresados. La validez entonces no es solo una cuestión de medida, la cual es regularmente esterilizada de un contexto, sino que representa la forma en que los resultados son legítimos para una realidad de los evaluados en particular. De esta manera, mientras más alejada sea la evaluación de un contexto, menos legítima puede ser para un conjunto diverso de realidades educativas que se viven de forma cotidiana en todo el país. Si bien una evaluación no retratará toda la *realidad* de cada participante, una propuesta comprensiva permitirá, al menos, un mayor acercamiento, así como valorar la

existencia de la complejidad del fenómeno aceptando las limitaciones que cualquier evaluación siempre tendrá.

En lo que respecta al análisis de los datos, la presente propuesta fomenta el uso de una medición basada en la norma. Esta visión de la medición ha sido abandonada en los últimos años al ser considerada como contraria a la objetividad, dado que se ha mantenido la idea que los puntajes deben ser estériles al contexto, resaltando la perspectiva que afirma que lo relevante es mantener un estándar inamovible para todos los evaluados. Bajo un esquema comprensivo, los estándares que si bien son inherentes en una retroalimentación, no pueden ser fijos ni ajenos al contexto y a la circunstancia bajo la cual se forma una persona, así como con el resto de las personas con las cuales se desarrolla la enseñanza. Esta forma de establecimiento de estándares tendrá que variar en cada aplicación que se desarrolle (la cual se sugiere sea de forma anual) y es que cada aplicación es única en sí y sus resultados deben ser tratados como tal.

Como parte del análisis también se debe considerar el cálculo de los psicométricos de los reactivos utilizados. El llamado *comportamiento de los reactivos* no depende de una respuesta que los estudiantes dan en *aislado*; es decir, las respuestas de los estudiantes nunca se dan por el conocimiento medido en sí mismo únicamente, sino que hay una variedad de factores que influyen en la respuesta que los chicos pueden dar a un reactivo. Como lo comenta Castro (2013), las respuestas a un reactivo pueden verse influenciadas incluso por las prácticas de enseñanza. En ese sentido, resulta importante no considerar los parámetros de los reactivos como unívocos o fijos, sino como una oportunidad de retroalimentar su diseño y promover su mejora para aplicaciones posteriores, con fines de perfeccionar su funcionamiento y buscar un progreso constante de los instrumentos utilizados.

El cierre de la propuesta de evaluación presentada se basa en la forma de comunicar los resultados. El caso analizado se enfocaba fuertemente en mostrar la proporción de egresados en diversos niveles de desempeño. Para el desarrollo de una propuesta comprensiva se debe privilegiar la retroalimentación más que la medición (normalmente representada por el puntaje de los egresados, independientemente de la escala utilizada), lo cual implica forzosamente enfocar la comunicación hacia los *stakeholders* interesados. Este cambio de visión debe considerar una comunicación más fluida y menos centrada en lo metodológico, privilegiando así la retroalimentación que se pueda dar a las IES y a los egresados, más que la comunicación de sus puntajes. Por ende, la retroalimentación debe tomar en cuenta forzosamente al contexto en el

cual se dan los puntajes obtenidos; de esta manera no se privilegian los datos en sí, sino lo relevante es la realidad bajo la cual se emiten. Esto es un serio problema que puede ser resuelto dese que se define el contexto y las variables en su análisis al momento de decidir los cuestionarios por utilizar, ya que estos elementos son, en sí, el sentido mismo de la evaluación.

Como regularmente se realiza en el caso de todas las propuestas de evaluación y de intervención en educación, es necesario considerar la posibilidad de implementar un ejercicio de evaluación bajo la perspectiva y metodología presentada. En particular de la evaluación, lo que la práctica indica es que la realidad misma determinará en gran medida la forma de llevar a cabo sus procesos. Por supuesto, cuando se trabaja bajo un modelo de servicio, algunas cuestiones y decisiones pueden modificarse dejando a un lado lo planeado y valorando la necesidad inmediata. Ningún trabajo de evaluación está exento de ello, sobre todo cuando hay elementos de política educativa y presupuesto involucrados; sin embargo, la instancia evaluadora deberá ser sensible a esto desde el momento de la planeación de la evaluación, dejando posibilidad de modificaciones durante el trayecto de la construcción de instrumentos, aplicación, análisis de datos y emisión de resultados, pero sin mermar la calidad de lo desarrollado y evitando desviar las decisiones hacia una perspectiva distinta a la comprensiva que fundamenta todas las actividades que implican un trabajo de este tipo.

Considerando el desarrollo de la propuesta presentada tal cual de la forma en que se plantea y con base en mi experiencia en la gestión de proyectos de evaluación a gran escala a nivel nacional e internacional, se requiere considerar al menos un año para el desarrollo y aplicación de los instrumentos que se consideren, adicional al tiempo de análisis de datos y emisión de resultados, lo cual en condiciones ideales de presupuesto y de número de colaboradores puede llevar al menos dos meses posteriores a la aplicación. Adicionalmente, de forma transversal a todo el proceso de evaluación se debe considerar elementos de presupuesto, los cuales suelen requerir millones de pesos, ello en función del amplio espectro de profesionistas que participan, viáticos, desarrollo de sistemas informáticos, papelería, imprenta y demás elementos que caracterizan a este tipo de evaluaciones.

Por otra parte, la viabilidad de este tipo de proyectos implica también valorar la posibilidad política de llevarlos a cabo. En este caso, las IES y sus programas educativos requieren de indicadores cuantitativos descriptivos precisos que refieran a la productividad alcanzada, medida tradicionalmente mediante la eficiencia terminal y de titulación. Como se ha comentado, los

resultados que los egresados obtienen en este tipo de evaluaciones son utilizados con fines predominantemente de certificación, así como un insumo para el financiamiento de los programas educativos. La presente propuesta no puede escapar de la demanda concreta que las instituciones realizan, sin embargo, ello no implica la renuncia al objetivo primordial de retroalimentación de la formación profesional de una manera profunda y contextualizada. Adicionalmente, como se ha comentado, este tipo de evaluación también resulta ser un elemento regulatorio del sistema educativo, mediante el control del egreso y titulación de miles de estudiantes cada año. Al respecto, los ejercicios de evaluación del egreso que actualmente se desarrollan en nuestro país resultan ser un sustituto del proceso de titulación de cada institución. En el caso de la propuesta realizada, ésta puede ser considerada como tal, en el sentido que representa una metodología más robusta que los ejercicios actuales y que puede ofrecer una mayor certeza de la validez de los resultados obtenidos para un egresado en particular, dada la consideración de diversos agentes en su elaboración, la medición de rasgos que no se centran únicamente en resultados de aprendizaje, así como la contextualización de las mediciones realizadas. Si bien no se propone establecer un estándar fijo o categorizar los resultados de los estudiantes en niveles de desempeño como lo hace Ceneval (Aún no satisfactorio, Satisfactorio y Sobresaliente), correspondería a cada programa educativo considerar la idoneidad esperada por sus egresados, el cual fungiría como estándar para la toma de decisión respecto a su titulación; de esta forma la prueba misma no funciona como un estándar en sí mismo.

La implementación de un proyecto de evaluación del egreso bajo la metodología propuesta considera entonces la experiencia que se ha tenido en México en este tipo de ejercicios, en particular la implementación de una cultura de evaluación dentro de las instituciones, así como los elementos operativos que implican. Adicionalmente, se debe considerar una relación estrecha con las IES respecto a hacer uso de una evaluación de este tipo bajo un enfoque comprensivo, resaltando las ventajas académicas que representa, sin mermar los usos e intereses administrativos que se han obtenido bajo los métodos tradicionales, los cuales no representan una discordancia con la estructura epistemológica y metodológica de este proyecto.

Finalmente, cabe resaltar la relevancia de la incorporación de diversos agentes dentro de todo el proceso del desarrollo de la evaluación. Es imposible que una instancia evaluadora cuente con todos los recursos y todas las personas que se requieren para una evaluación de tal magnitud. Por tanto, es necesaria la consideración de diversos agentes no solo por una cuestión operativa y de

fluidez de procesos laborales, ya que entendiendo a la evaluación como la construcción del marco de una realidad bajo la cual ciertas personas serán comparadas (aún sea con fines de retroalimentación), el mayor soporte que se debe tener no es solo una coherencia epistemológica, sino un sustento de diversas perspectivas respecto al mismo fenómeno de interés. La definición del marco de evaluación representa una especie de fragmentación y parcelación de una posibilidad de posturas –todas de ellas válidas en cierta medida– sobre lo que interesa retroalimentar. No será el mismo resultado una evaluación cuya constitución teórica sea desarrollada bajo la consideración de una revisión de literatura sobre el fenómeno de interés, que aquella que se construya mediante la opinión de un grupo de personas. Ambas perspectivas resaltarán diversos aspectos de interés que apoyarán la validez de la emisión de los resultados bajo una tendencia en particular definida por los intereses mismos de la fuente de información utilizada. Por supuesto, el uso de elementos teóricos da validez académica a costa del involucramiento social de los productos que se generen, mientras que el uso de opiniones favorece la contextualización en la interpretación de resultados pero regularmente es considerada con poca seriedad dentro del ámbito académico debido a su relativa falta de objetividad. Aunque partiendo del hecho que la realidad de un fenómeno educativo no es en sí misma, sino en función de los actores y medios involucrados, la postura comprensiva que da sostén a la presente propuesta realza la diversidad de posturas frente a un mismo fenómeno de interés, de tal suerte que la conjunción de esta diversidad permita la construcción de un modelo unificador. Como consecuencia, la relativa objetividad siempre privilegiada por las metodologías de evaluación a gran escala se observa de otra forma; bajo esta postura, la objetividad en realidad se relaciona con la convicción de que el conocimiento producido encuentre en una comunidad determinada un suficiente grado de aceptación, lo cual a su vez da mayor sentido a los resultados, privilegiando también su forma de comunicación, así como enfocando su uso hacia la retroalimentación de la formación profesional de una persona en particular.

## Referencias

- Acosta, A. (1995). Políticas públicas de educación superior y universidades en México, 1982-1992. *Estudios sociológicos*, 13(38).
- ADB. (2005). *Project performance audit report. Higher education project in Indonesia*.
- Alvarado, H. y Batanero, C. (2006). El significado del teorema central del límite: evolución histórica a partir de sus campos de problemas. En: Contreras, A., Ordoñez, L. y Batanero, C. (Eds.). *Investigación en didáctica de las matemáticas / Congreso internacional sobre aplicaciones y desarrollos de la teoría de las funciones semióticas*. 257 – 277.
- Álvarez, J. (2001). *Evaluar para conocer, examinar para excluir*. Madrid: Morata.
- Anderson, J. (2005). *La rendición de cuentas en la educación*. Instituto Internacional para la Planeación de la Educación: Paris.
- Antúnez, A. y Aranguren, C. (1998). Aproximación teórica y epistemológica al problema de la evaluación. Su condición en educación básica. *Boletín Teoría y didáctica de las ciencias sociales*, 3(2). 104-117.
- Antúnez, A. y Aranguren, C. (2004). Problemática teórico-filosófica de la evaluación en la educación básica durante las dos últimas décadas del Siglo XX. *EDUCERE*, 8(25). 149-153.
- Bolívar, A. (2005). El lugar de la ética profesional en la formación universitaria. *Revista Mexicana de Investigación Educativa*, 10(24). 93-123.
- Bonafé, J. (1988). El estudio de casos en la investigación educativa. *Investigación en la Escuela*, 6. 41 – 50.
- Brower, J. (2009). Fundamentos epistemológicos para el esbozo de una pedagogía completa. *Revista de la Universidad Bolivariana*, 8. 57 – 86.
- Calero, J. y Choi, A. (2012). La evaluación como instrumento de política educativa. *Presupuesto y gasto público*, 67. 29 – 42.
- Camacho, C. y Sánchez, F. (1997). *Psicometría*. Sevilla: Kronos.
- Cano, M. (2008). La evaluación por competencias en la educación superior. *Profesorado. Revista de currículum y formación del profesorado*, 12(3).

Casanova, M. (1998). Evaluación: concepto, tipología y objetivos. *Biblioteca para la actualización del maestro*. 67-101.

Casillas, M. y De Garay, A. (1993). Lo público y lo privado en la educación superior en México. *Sociológica*, 8(22).

Castro, D. (2013). Prueba de PISA 2009. La competencia lectora. Una explicación didáctica sobre las posibles causas del bajo nivel de desempeño de los estudiantes mexicanos en la subescala “Integrar e interpretar”. *Tendencias pedagógicas*, 21. 199 – 213.

Ceneval (2012a). Análisis de resultados del EGEL e ingreso al Padrón de Programas de Alto Rendimiento (IDAP). Presentación en la Universidad Veracruzana del 5 de marzo de 2012. Consultado el 11 de noviembre de 2014 en: [www.uv.mx/planeacioninstitucional/files/.../PresentacionEGELUV.pptx](http://www.uv.mx/planeacioninstitucional/files/.../PresentacionEGELUV.pptx)

Ceneval (2012b). Se entregó por primera vez el Premio Ceneval, un reconocimiento a la excelencia. Nota consultada el 8 de diciembre de 2014 en: <http://www.ceneval.edu.mx/ceneval-web/content.do?page=6679>

Ceneval (2013a). *Cuarta convocatoria para el ingreso al Padrón de Programas de Licenciatura de Alto Rendimiento Académico*. Consultado el 04 de noviembre del 2014 en: [http://archivos.ceneval.edu.mx/documentos/idap/Convocatoria\\_2013\\_2014.pdf](http://archivos.ceneval.edu.mx/documentos/idap/Convocatoria_2013_2014.pdf)

Ceneval (2013b). *Manual técnico del EGEL*. México: Ceneval.

Ceneval (2013c). *Metodología Ceneval*. México: Ceneval.

Ceneval (2015a). *Resultados del Examen Nacional de Ingreso a la Educación Media Superior en el año 2014. EXANI-I*. México: Ceneval

Ceneval (2015b). *Resultados del Examen Nacional de Ingreso a la Educación Media Superior en el año 2014. EXANI-II*. México: Ceneval

Ceneval (2015c). *Resultados del Examen Nacional de Ingreso a la Educación Media Superior en el año 2014. EXANI-III*. México: Ceneval

Cepeda, B. (2009). La política de evaluación externa de los aprendizajes en educación básica. Una aproximación a la comprensión de su racionalidad. Periodo 2000 – 2006. *Tesis doctoral no publicada*. México: UNAM

- Chaín, R., Cruz, N., Martínez, M., Jácome, N. (2003). Examen de selección y trayectoria escolar. *Revista de la educación superior*, 125(32).
- Colom, A. (2005). Teoría del caos y práctica educativa. *Revista galega do ensino*, 47.
- Contreras, O. (s/f). La educación como objeto de estudio de las ciencias sociales. *Unidad Mínima de Aprendizaje del seminario "Historia, Sociedad y Educación" de la Maestría en Docencia para la Educación Media Superior*. UNAM.
- Cortada de Kohan, C. (2004). Teoría de Respuesta al Ítem: supuestos básicos. *Evaluar*, 4.
- Cortada, N. (1999). *Teorías psicométricas y construcción de tests*. Buenos Aires: Lugar.
- Cruz, Y. y Cruz, A. (2008). La educación superior en México. Tendencias y desafíos. *Avaliação: Revista da Avaliação da Educação Superior*, 13(2).
- Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3). 297-334.
- De la Cruz, G.; Leyva, Y. y Seda, I. (2013). La evaluación del aprendizaje: Tendencias teóricas y metodológicas en el periodo 2001 – 2010. En: Maldonado, A. (Coordinador). *Educación y ciencia: políticas y producción del conocimiento 2002-2011*. Colección Estados del Conocimiento. México: ANUIES-COMIE.
- De la Garza, J. (2008). Evaluación y acreditación de la educación superior en América Latina y el Caribe. En Tünnermann (Ed.). *La educación superior en América Latina y el Caribe: diez años después de la Conferencia Mundial de 1998*. Cali: Sello Editorial Javeriano. 175-222.
- DePascale, Ch. (2003). The ideal role of large-scale testing in a comprehensive assessment system. *Association of Test Publishers*, 5(1). 1-11.
- Díaz Barriga, A. (1994). Una polémica en relación al examen. *Revista Iberoamericana de Educación*, 5. 161–181
- Domínguez, S. (2013). ¿Ítems politómicos o dicotómicos? Un estudio empírico con una escala unidimensional. *Revista argentina de ciencias del comportamiento*, 5(3). 30-37.
- Evaluation of the Noth Lanarkshire curriculum flexibility project (2004). *Quality in Education Centre University of Strathclyde*.
- Fernández, A. (2010). La evaluación orientada al aprendizaje en un modelo de formación por competencias en la educación universitaria. *Revista de docencia universitaria*, 8(1). 11-34.

Fernández, J. (2001). Elementos que consolidan el concepto de profesión. Notas para su reflexión. *Revista Electrónica de Investigación Educativa*, 3(1). 23-39.

Final report of the external evaluation of the Charter School of Education. California State University, Los Angeles. (2000). *The Evaluation Center. Western Michigan University*.

González, V. (2002). ¿Qué significa ser un profesional competente? Reflexiones desde una perspectiva psicológica *Revista Cubana de Educación Superior*, XXII(1). 45-53.

González, V. (2006). La formación de competencias profesionales en la universidad. Reflexiones y experiencias desde una perspectiva educativa. *Revista de educación*. 175-187.

González, V. y González, R. (2008). Competencias genéricas y formación profesional: un análisis desde la docencia universitaria. *Revista Iberoamericana de Educación*, 47. 185-209.

González-Montesinos, M. y Backhoff, E. (2010). Validación de un cuestionario de contexto para evaluar sistemas educativos con modelos de ecuaciones estructurales. *Revista electrónica de investigación y evaluación educativa*, 16(2). 1 – 17.

Hambleton, R., Jaeger, R., Plake, B., Mills, C. (2000). Setting performance standards on complex educational assessments. *Applied psychological measurement*, 24. 355-366.

Heredia, A. (2009). Dos formas diferenciadas de evaluación didáctica: evaluación normativa para seleccionar a los alumnos y evaluación criterial para el dominio del conocimiento básico. *Bordón*, (4). 39-48.

Hernández (2011). *XIII Reunión Ordinaria de la Asociación Nacional de Escuelas y Facultades de Pedagogía y Educación, A.C. (ANEFEP) del 30 de septiembre de 2011*. Consultado el día 03 de noviembre de 2014 en: [anefep.org.mx/wp-content/uploads/2011/10/IDAP.pptx](http://anefep.org.mx/wp-content/uploads/2011/10/IDAP.pptx)

Hernández, J., Márquez, A. y Palomar, J. (2006). Factores asociados con el desempeño académico en el EXANI-I. *Revista Mexicana de Investigación Educativa*, 11(29). 547–581.

Hernández, A. y Rodríguez, K. (2008). La Organización para la Cooperación y el Desarrollo Económico, OCDE, y la Definición de Competencias en Educación Superior: el caso de México. *Educere*, 12(43). 751-758.

Hontagas, P., Ponsoda, V., Olea, J. y Revuelta, J. (1998). ANATRI 1.0: Representación de funciones características de ítems dicotómicos y politómicos. *Psicothema*, 10(2). 475–479.

Horbath, J. y Gracia, M. (2014). La evaluación educativa en México. *Revista de relaciones internacionales, estrategia y seguridad*, 9(1). 59–85

INEE (2006). *Plan Maestro de desarrollo 2007 – 2014*. México: INEE.

INEE (2014). *El derecho a una educación de calidad*. Informe 2014. México: INEE.

INEE (2015). *Las pruebas Excale y Enlace. Un estudio de validación*. México: INEE.

ITESM (2013). *Tecnológico de Monterrey: líder en calidad académica en Premios Ceneval*. Nota publicada el 2 de diciembre de 2013 en su portal. Consultada el 4 de noviembre de 2014 en: <http://www.itesm.mx/wps/wcm/connect/snc/portal+informativo/por+tema/educacion/entregare+conocimientosceneval2dic13>

Jiménez, J., García, J. y González, F. (2006). Guía técnica para la construcción de cuestionarios. *Odiseo. Revista electrónica de pedagogía*, 3(6).

Jornet, J., González, J. y Perales, J. (2012). Diseño de cuestionarios de contexto para la evaluación de sistemas educativos: optimización de la medida de constructos complejos. *Bordón*, 64(2). 89–110.

Karantonis, A. y Sireci, S. (2006). The bookmark standard-setting method: A literature review. *Educational measurement: Issues and practice*, 25(1). 4-12.

Kuhn C., Zlatkin-Troitschanaskaia, O. & Bergsman, K. (2011). *Assessment of Competencies among University Students and Graduates: Analyzing the state of Research and Perspectives*. Johannes Gutenberg Universität Mainz.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22(140).

Lin, J. (2006). The bookmark procedure for setting cut-scores and finalizing performance standards: strengths and weaknesses. *The Alberta Journal of Educational Research*, 52(1). 36-52.

López, M. (2013). Ética profesional y complejidad. Los principios y la religación. *Perfiles educativos. Suplemento 2013. Ética profesional en la educación superior*, 35(142). 43–52.

Martínez Arias, R. (2006). La metodología de los estudios PISA. *Revista de educación. Extraordinario*. 111–129.

Martínez, F. (2001). Evaluación educativa y pruebas estandarizadas. Elementos para enriquecer el debate. *Revista de la educación superior*, 30(4). 71-85.

- Martínez, F. (2003). *Pruebas y rendición de cuentas*. México: INEE
- Martínez, F. (2008a). *Evaluación a gran escala y evaluación en el aula*. Los usos de Excale para la mejora educativa. Consultado el 1 de noviembre de 2013 en: <http://www.inee.edu.mx/explorador/propuesta2008/Textos/EvaluacionGranEscala.pdf>
- Martínez, F. (2008b). *La evaluación de aprendizajes en América Latina*. México: INEE.
- Martínez, F. (2009). Evaluación formativa en el aula y evaluación a gran escala; hacia un sistema más equilibrado. *Revista electrónica de investigación educativa*, 11(2).
- Martínez, F. y Santos, A. (2009). Evaluación e investigación educativa. En: Alba, A. (Coord.). *¿Qué dice la investigación educativa?* México: Consejo Mexicano de Investigación Educativa. 265-304.
- Martínez, F., Cepeda, B., Fuentes, G. y Burgos, R. (1995). Evaluación del aprendizaje. En: Rueda, M (Coord.). *La investigación educativa en los ochenta, perspectivas para los noventa. Procesos de enseñanza y aprendizaje I*. México: Consejo Mexicano de Investigación Educativa.
- Martínez, S. (1998). El objeto de la sistematización y sus relaciones con la evaluación y la investigación. *Ponencia presentada en el Seminario latinoamericano: sistematización de prácticas de animación sociocultural y participación ciudadana en América Latina. Medellín, Fundación Universitaria Luis Amigó-CEAAL*. 12-14 de agosto.
- Márquez, A. (2004). Calidad de la educación superior en México. ¿Es posible un sistema eficaz, eficiente y equitativo? Las políticas de financiamiento de la educación superior en los noventa. *Revista Mexicana de Investigación Educativa*, 9(21). 477-500.
- Meckes, L. (2007). Evaluación y estándares: logros y desafíos para incrementar el impacto en calidad educativa. *Pensamiento educativo*, 40(1). 351-371.
- Mertler, Craig A. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research & Evaluation*, 7(25).
- Morin, E. y Pakman, M. (1994). El paradigma de la complejidad. En: *Introducción al pensamiento complejo*. Barcelona: Gedisa.
- Morin, E. (2004). La epistemología de la complejidad. *Gazeta de antropología*, 20.
- Moreno, T. (2009). Competencias en educación superior: un alto en el camino para revisar la ruta de viaje. *Perfiles educativos*, 31(124).

- Moreno, T. (2010). Lo bueno, lo malo y lo feo: las muchas caras de la evaluación. *Revista Iberoamericana de Educación Superior*, 1(2).
- Morrison, K. (2003) Complexity theory and curriculum reforms in Hong Kong. *Pedagogy, Culture & Society*, 11(2). 279-302.
- Muñiz, J. (1997). *Introducción a la Teoría de Respuesta a los Ítems*. Madrid: Pirámide.
- Muñiz, J. (1998). La medición de lo psicológico. *Psicothema*, 10(1). 1-21.
- Muñiz, J., Fidalgo, A., Cueto, E., Martínez, R. y Moreno, R. (2005). *Análisis de los ítems*. Madrid: Muralla.
- Murillo, J. (2008). Los modelos multinivel como herramienta para la investigación educativa. *Magis. Revista Internacional de Investigación en Educación*, 1. 45–62.
- Nevo, B. (1985). Face validity revisited. *Journal of educational measurement*, 22(4). 287–293.
- Nicol, E. (2013). *Psicología de las situaciones vitales*. México: Fondo de cultura económica.
- Norcini, J. (2003). Setting standards on educational tests. *Medical education*, 37. 464-469.
- NRF. (2001). Report on the evaluation of the NRF programmes directed at research and RESEARCH CAPACITY development at historical black universities (HBUs).
- NSF/EH/US.DE. Postsecondary Education Institution's Satisfaction with Student Financial Assistance Programs.
- Nusche, D. (2008). Assessment of Learning Outcomes in Higher Education: a comparative review of selected practices. *OECD education Working Papers*. No. 15, OECD Publishing . Consultado el 8 de noviembre de 2013 en: <http://dx.doi.org/10.1787/244257272573>
- OCDE (2015a). *PISA 2015 Draft questionnaire framework*. Paris: OCDE.
- OCDE (2015b). *PISA 2015 released field trial cognitive items*. Consultado el 18 de marzo de 2015 en: <http://www.oecd.org/pisa/pisaproducts/PISA2015-Released-FT-Cognitive-Items.pdf>
- Oliver, S. y Hyun, E. (2008). Comprehensive curriculum reform in higher education: collaborative engagement of faculty and administrators. *Journal of case studies in education*, 2(1).
- Peña, W. (2009). El estudio de caso como recurso metodológico apropiado para la investigación en ciencias sociales. *Educación y desarrollo social*, 2(3). 180–195.

Perales, M. y Pérez, A. (2009). En: Jornet, J. y Leyva, Y. (Comp). *Conceptos, metodología y profesionalización de la evaluación educativa*. México: INITE.

Pereira, Z. (2011). Los diseños de método mixto en la investigación en educación: una experiencia concreta. *Revista Electrónica Educare, XV(1)*. 15-29.

Pérez, E. y Tornibeni, S. (2004). Construcción de tests. En: Tornimbeni, S., Pérez, E., Olaz, F. y Fernández, A. (Coord.). *Introducción a los tests psicológicos*. Córdoba: Brujas.

Pinuel, J. (2002). Epistemología, metodología y análisis de contenido. *Estudios de sociolingüística, 3(1)*. 1–42.

Ramos, G., Perales, M. y Pérez, A. (2009). En: Jornet, J. y Leyva, Y. (Comp.). *Conceptos, metodología y profesionalización de la evaluación educativa*. México: INITE.

Ravela, P. (2006). Para comprender las evaluaciones educativas. Fichas didácticas. Programa de promoción de la Reforma Educativa en América Latina y el Caribe. Consultado el 23 de octubre de 2013

en:

[http://www.iae.org.mx/2012/Dip1DGIFA/nov1/S102\\_Ravela\\_Pedro\\_Fichas\\_didacticas.pdf](http://www.iae.org.mx/2012/Dip1DGIFA/nov1/S102_Ravela_Pedro_Fichas_didacticas.pdf)

Rychen, D. y Salganik, L. (2003, abril). Highlights from the OECD Project Definition and Selection Competences: Theoretical and Conceptual Foundations. *Presentación en el Annual Meeting of the American Educational Research Association*. Chicago.

Rodríguez, R. (2002). Continuidad y cambio de las políticas de educación superior. *Revista Mexicana de Investigación Educativa, 7(14)*. 133–154.

Rodríguez, R. (2003). La educación superior en el mercado: configuraciones emergentes y nuevos proveedores. En: Mollis, M. (Coomp.). *Las universidades en América Latina: ¿Reformas o alteradas? La cosmética del poder financiero*. Buenos Aires: Consejo Latinoamericano de Ciencias Sociales.

Romero, A. (2012). Propuesta de evaluación para las competencias cívicas y éticas para la educación primaria en México. *Tesis de maestría no publicada*. México: UNAM.

Rueda, M. (2008). La evaluación del desempeño docente en las universidades públicas de México. *Revista Iberoamericana de Evaluación Educativa, 1(3)*.

Ruiz, A., Pardo, A. y San Martín, R. (2010). Modelos de ecuaciones estructurales. *Papeles del psicólogo, 31(1)*. 34 – 45.

- Ruiz, C. (2008). El enfoque multimétodo en la investigación social y educativa: una mirada desde el paradigma de la complejidad. *Revista de filosofía y sociopolítica de la educación*, 8(4). 13-28.
- Salinas, D. (2002). *Mañana examen. La evaluación: entre la teoría y la realidad*. Barcelona: Grao.
- Santos Guerra, M. (1988). Patología general de la evaluación educativa. *Infancia y aprendizaje*, 41. 143-158.
- Schutz, A. (1974). *El problema de la realidad social*. Buenos Aires: Amorroutu
- SEP (2002). *La calidad de la educación en México: Perspectivas, análisis y evaluación*. México: Porrúa.
- SEP (2007). *Manual técnico ENLACE*. México: SEP.
- SEP (2014). *Programa Integral de Fortalecimiento Institucional (PIFI)*. Página web consultada el día 11 de noviembre de 2014 en: <http://pifi.sep.gob.mx/>
- Sierra, F., Valdelamar, J., Hernández, F. y Sarmiento, L. (2007). Test adaptativos informatizados. *Avances en medición*, 5. 157–162.
- Stake, R. (2005). *Investigación con estudios de caso*. Madrid: Morata.
- Stake, R. (2006). *Evaluación comprensiva y evaluación basada en estándares*. Barcelona: Graó.
- Stufflebeam, D. y Shinkfield, A. (1993). *Evaluación sistemática. Guía teórica y práctica*. Madrid: Paidós.
- Tiana, A. (2012a). Analizar el contexto para obtener el máximo beneficio de la evaluación. *Bordón*, 64(2). 15-28.
- Tiana, A. (2012b). Calidad, evaluación y estándares: algunas lecciones de las reformas recientes. En: Marchesi, A., Tedesco, J. y Coll. C. (Coord.). *Calidad, equidad y reformas en la enseñanza*. Madrid: OEI. 113-124.
- Tristán, A. (2012). *Perfil de los lectores de los estudiantes de secundaria de PISA*. México: INEE.
- Tuirán, R. (2011). La educación superior en México: avances, rezagos y retos. *Educación contracorriente*.
- UABC (2013). *Acapara UABC calidad académica*. Nota publicada el 22 de agosto de 2013 en su portal. Consultada el 11 de noviembre de 2014 en:

<http://www.frontera.info/EdicionEnLinea/Notas/Noticias/22082013/744133-Acapara-UABC-calidad-academica.html>

UAEH (2014). *Padrón de programas de licenciatura de alto rendimiento EGEL*. Consultado el 11 de noviembre de 2014 en:

[http://www.uaeh.edu.mx/padron\\_programas\\_licenciaturas\\_alto\\_rendimiento\\_egel/](http://www.uaeh.edu.mx/padron_programas_licenciaturas_alto_rendimiento_egel/)

UAEM (2013). *Premio Ceneval al desempeño de excelencia-EGEL*. Consultado el 4 de noviembre de 2014 en: [www.facico-uaemex.mx/descargas/egel.pdf](http://www.facico-uaemex.mx/descargas/egel.pdf)

UNESCO (2005). *Hacia las sociedades del conocimiento*. UNESCO: 2005.

US Department of Education. *Evaluation of the public charter schools program: Year one evaluation report*. Washington, D.C.

US Department of Education. *When schools stay open late: The national evaluation of the 21 st century community learning centers program*.

Villa, A. y Poblete, M. (2011). Evaluación de competencias genéricas: principios, oportunidades y limitaciones. *Bordón*, 63(1). 147-170.

Villaseñor, G. (2003). La evaluación de la educación superior: su función social. *Reencuentro*, 36. 20-29.

Viñao, A. (2012). El desmantelamiento del derecho a la educación: discursos y estrategias neoconservadoras. *Revista Internacional de Ciencias Sociales*, 31. 97-107.

Volante, L. (2006). An alternative vision of large scale assessment in Canada. *Journal of teaching and learning*, 4(1).