

Concepto y funciones de la Investigación Evaluativa y de la Evaluación de Programas Educativos

1

ESQUEMA/CONTENIDOS

INTRODUCCIÓN

1.1. LA EVALUACIÓN DE PROGRAMAS Y LA INVESTIGACIÓN EVALUATIVA

- 1.1.1. Orígenes y consolidación en el ámbito internacional
- 1.1.2. La pluralidad metodológica
- 1.1.3. Generaciones de la evaluación: de la medición a la construcción social
- 1.1.4. La investigación evaluativa como disciplina
- 1.1.5. La evaluación en el ámbito iberoamericano y español
- 1.1.6. Retos actuales de la investigación evaluativa

1.2. EL CONCEPTO DE EVALUACIÓN DE PROGRAMAS

- 1.2.1. Definición y evolución conceptual
- 1.2.2. Dimensiones comprensivas del proceso evaluativo
- 1.2.3. Evaluabilidad como requisito previo
- 1.2.4. La dualidad teleológica: mejora y rendición de cuentas
- 1.2.5. Sistema comprensivo de criterios evaluativos
- 1.2.6. Aplicaciones prácticas y estudios de caso
- 1.2.7. Retos contemporáneos y perspectivas futuras

1.3. EL CONCEPTO DE CALIDAD. CRITERIOS Y REFERENCIAS PARA LA EVALUACIÓN DE LA CALIDAD DE LOS PROGRAMAS

- 1.3.1. La polisemia del concepto de calidad
- 1.3.2. Calidad y educación: un concepto en disputa
- 1.3.3. Criterios de calidad en programas educativos
- 1.3.4. Tendencias actuales en los sistemas de evaluación y acreditación
- 1.3.5. Estándares internacionales de calidad
- 1.3.6. Tensiones, críticas y retos contemporáneos

1.3.7. Ejemplos prácticos de evaluación de la calidad

1.4. LAS FUNCIONES DE LA EVALUACIÓN DE PROGRAMAS. PROPÓSITOS DE LA EVALUACIÓN DE PROGRAMAS

1.4.1. La Dualidad Funcional Fundamental: Formativa y Sumativa

1.4.2. Funciones técnicas, políticas y sociales

1.4.3. Propósitos Centrales de la Evaluación de Programas

1.4.4. La Evaluación como Competencia Profesional y Motor de Aprendizaje Organizacional

1.4.5. Evaluación y Responsabilidad Social

1.4.6. Funciones de la evaluación en el ciclo de vida de los programas

1.4.7. Retos actuales en torno a las funciones de la evaluación

1.4.8. Ejemplos ilustrativos

SÍNTESIS CAPÍTULO 1

REFERENCIAS BIBLIOGRÁFICAS

INTRODUCCIÓN

La evaluación educativa ocupa hoy un lugar central en el análisis y la gestión de los sistemas escolares, ya que no constituye únicamente una herramienta técnica, sino también un espacio de debate social y político en el que se dirimen cuestiones de legitimidad, confianza pública y rendición de cuentas. Como señala Richardson (2022), la evaluación es, al mismo tiempo, un discurso público que configura las percepciones sociales sobre la educación. Por ello, no puede pensarse exclusivamente desde la validez técnica de sus procedimientos, sino también desde su aceptación social y su capacidad para generar confianza en las instituciones educativas.

En su sentido más general, la evaluación se concibe como un proceso valorativo orientado a determinar el mérito, la calidad y la utilidad de una persona, actividad o programa (Scriven, 2001). En el ámbito educativo, esta definición se amplía y complejiza, al convertirse en un dispositivo estratégico que opera en múltiples niveles: desde el aula, mediante la valoración del aprendizaje del alumnado, hasta la política educativa, a través de la evaluación de programas, instituciones y sistemas en su conjunto.

Los fines de la evaluación educativa han sido objeto de debate permanente. Tradicionalmente, se ha distinguido entre evaluación formativa, orientada a retroalimentar los procesos de enseñanza y aprendizaje para mejorarlos, y evaluación sumativa, destinada a certificar resultados, rendir cuentas o apoyar la toma de decisiones administrativas (Black & Wiliam, 1998). Ambas funciones resultan necesarias, aunque su integración no está exenta de tensiones, ya que las exigencias externas de control pueden entrar en conflicto con la función pedagógica de mejora. En este sentido, autores como Wiliam (2011, 2018) han subrayado que la evaluación debe concebirse prioritariamente como una herramienta para el aprendizaje, destacando el papel de la retroalimentación efectiva como uno de los factores con mayor impacto en el rendimiento del alumnado.

Desde una perspectiva metodológica, la evaluación se ha desarrollado en torno a dos grandes paradigmas. El primero, de carácter positivista, privilegia la medición cuantitativa, las pruebas estandarizadas y los diseños experimentales con el objetivo de garantizar la fiabilidad y la comparabilidad de los resultados. El segundo, de orien-

tación interpretativa o fenomenológica, pone el acento en los métodos cualitativos, los estudios de caso y la observación en contextos naturales, buscando comprender los significados y los procesos educativos. Aunque durante décadas ambos enfoques se presentaron como opuestos, la evolución del campo ha evidenciado la necesidad de su complementariedad. En la actualidad, los enfoques mixtos y la triangulación metodológica permiten obtener una comprensión más rica, contextualizada y útil de los fenómenos educativos (Erickson, 1989; Herman, 2013; LeCompte, 1995; Wiliam, 2018).

En las últimas décadas, la evaluación ha ampliado notablemente su campo de acción. Ya no se limita al ámbito del aula, sino que se extiende al análisis del desempeño docente, la gestión institucional y, de forma creciente, a la evaluación de programas y políticas educativas de gran escala (Casassus, 2000; De la Garza, 2004). Este proceso ha contribuido a consolidarla como una herramienta clave de gobernanza educativa, estrechamente vinculada a la asignación de recursos, la definición de estándares y la legitimación de políticas públicas. Paralelamente, investigaciones recientes muestran que la evaluación no solo mide aprendizajes, sino que también influye en la motivación del alumnado, en su autopercepción de competencia y en su compromiso con el proceso formativo (Martínez-Garrido & Murillo, 2023). De este modo, la evaluación se configura como un proceso con implicaciones pedagógicas, psicológicas y sociales que trascienden su dimensión estrictamente técnica.

Los retos contemporáneos de la evaluación se intensifican con la irrupción de la tecnología y la consolidación de la sociedad del conocimiento. En este nuevo escenario, la evaluación incorpora progresivamente competencias no cognitivas –como la creatividad, la colaboración o la comunicación– y se apoya en modelos multimodales potenciados por la inteligencia artificial (Ying et al., 2024). Sin embargo, estas innovaciones plantean importantes dilemas éticos relacionados con el sesgo algorítmico, la transparencia de los sistemas y el uso responsable de los datos educativos (Ying et al., 2024). Todo ello refuerza la idea de que la evaluación no es un proceso neutro, sino una práctica social situada que exige integrar de manera explícita principios éticos y de justicia social en su diseño, desarrollo e interpretación.

En este contexto amplio y complejo se sitúa el presente tema, centrado en el análisis del concepto y las funciones de la investigación evaluativa y de la evaluación de programas. Su estudio permite al alumnado adentrarse en los fundamentos históricos, epistemológicos y conceptuales de este campo, así como en los distintos enfoques y criterios de calidad que lo orientan. A través de este recorrido, se sientan las bases necesarias para comprender y aplicar posteriormente los modelos, métodos y estándares que estructuran la práctica evaluativa contemporánea.

Desde una perspectiva formativa, el desarrollo de este tema contribuye a consolidar competencias fundamentales en el ámbito de la evaluación educativa. En particular, permite desarrollar una comprensión detallada de los fundamentos teóricos que sustentan la evaluación, potenciar la capacidad de interpretar críticamente informes de investigación evaluativa en función de su evolución histórica y metodológica, y for-

talecer la competencia para formular juicios fundamentados sobre la pertinencia, la calidad y las funciones de las evaluaciones en contextos educativos reales.

De manera más específica, el trabajo realizado en este tema facilita que el estudiante comprenda los fundamentos históricos, epistemológicos y conceptuales de la investigación evaluativa; distinga entre investigación social, investigación educativa y evaluación de programas; identifique criterios de calidad en programas e instituciones; y reflexione críticamente sobre las funciones sociales y políticas de la evaluación, en particular en lo relativo a la rendición de cuentas, el aprendizaje organizacional y la mejora. Asimismo, permite aplicar el sistema comprehensivo de criterios de calidad –relevancia, eficacia, eficiencia, impacto, sostenibilidad y equidad– al análisis de programas educativos, así como analizar las tensiones contemporáneas entre las funciones formativas y sumativas en contextos específicos.

En síntesis, este tema no solo introduce los conceptos fundamentales de la investigación evaluativa, sino que también establece un marco de comprensión que integra teoría, método y práctica, preparando al estudiante para abordar la evaluación de programas como un proceso riguroso, contextualizado y orientado a la mejora y a la toma de decisiones fundamentadas en la educación contemporánea.

1.1. LA EVALUACIÓN DE PROGRAMAS Y LA INVESTIGACIÓN EVALUATIVA

La investigación evaluativa se ha consolidado como una de las áreas más dinámicas y necesarias dentro de las ciencias sociales y de la educación. A nivel epistemológico y operativo, resulta pertinente establecer una distinción fundamental entre los términos «investigación evaluativa» y «evaluación de programas», pese a su frecuente uso coloquial como sinónimos. Esta delimitación enriquece la comprensión del campo de estudio.

La investigación evaluativa puede ser conceptualizada como la dimensión metodológico-disciplinar que provee el marco teórico, los principios y los procedimientos sistemáticos para el estudio de programas, políticas o intervenciones. Se erige como una modalidad de investigación aplicada cuyo propósito último es la generación de conocimiento fundamentado y basado en evidencia para la toma de decisiones (Stufflebeam & Shinkfield, 2002). Se caracteriza por su naturaleza cíclica, la integración de paradigmas metodológicos cuantitativos y cualitativos, y su orientación hacia la utilidad para las partes interesadas (*stakeholders*).

Por contraste, la evaluación de programas representa la dimensión práctica-aplicada. Es decir, la implementación concreta de los principios de la investigación evaluativa para emitir un juicio fundamentado sobre el mérito y el valor de una intervención específica en un contexto determinado. Constituye una práctica organizacional esencial en ámbitos como la educación o la salud pública, centrándose en la comprensión,

el perfeccionamiento y la rendición de cuentas de iniciativas diseñadas para modificar condiciones en poblaciones objetivo.

En síntesis, mientras la investigación evaluativa constituye el *corpus* de conocimiento que prescribe «cómo» y «por qué» evaluar, la evaluación de programas es la praxis que materializa dicho conocimiento. Ambas dimensiones son consustanciales e interdependientes, conformando un continuo entre la teoría y la práctica. A diferencia de otras formas de indagación científica, no se limitan a describir o explicar fenómenos, sino que se centran en valorar la calidad, pertinencia, eficacia e impacto de programas e intervenciones sociales y educativas. Esta finalidad práctica no excluye el rigor científico; al contrario, exige una metodología sistemática que permita obtener información fiable, válida y útil para fundamentar la toma de decisiones.

1.1.1. Orígenes y consolidación en el ámbito internacional

La emergencia de la investigación evaluativa se vincula históricamente al contexto estadounidense de los años sesenta, marcado por la *War on Poverty* y por el crecimiento de programas sociales a gran escala. En este escenario, la administración pública necesitaba disponer de información rigurosa para justificar inversiones, rendir cuentas a la ciudadanía y orientar políticas públicas. Howard Suchman (1967) fue uno de los pioneros en definir la investigación evaluativa como la aplicación de métodos de investigación social para determinar la efectividad de programas.

Sin embargo, los primeros enfoques se caracterizaron por un excesivo experimentalismo. Inspirados en el modelo de las ciencias naturales, se privilegiaban diseños experimentales y cuasi-experimentales orientados a establecer relaciones de causalidad entre la intervención y los resultados. Esta perspectiva fue criticada por autores como Lee Cronbach (1982), quien señaló que los programas sociales y educativos no pueden evaluarse con las mismas reglas que los experimentos de laboratorio. Para Cronbach, lo esencial no es solo medir efectos, sino entender cómo y por qué un programa funciona en determinados contextos, de modo que la evaluación aporte orientaciones útiles para la práctica.

A partir de estas críticas, la investigación evaluativa comenzó a ampliar su horizonte, incorporando nuevos enfoques metodológicos y epistemológicos. Ralph Tyler, ya en los años cuarenta, había introducido la idea de que la evaluación debía entenderse como el proceso de comprobar el grado en que se alcanzan los objetivos educativos. Posteriormente, Michael Scriven (1967, 1991) propuso que la evaluación no se limitaba a verificar objetivos, sino que consistía en un proceso autónomo de juicio de valor, con dos funciones básicas: la formativa, orientada a mejorar durante el desarrollo del programa, y la sumativa, destinada a emitir un veredicto final.

En paralelo, Guba y Lincoln (1989) plantearon una cuarta generación de evaluación que se alejaba del positivismo clásico y defendía un enfoque construccionista y

participativo. Desde su perspectiva, la evaluación es un proceso de negociación en el que los distintos actores implicados construyen colectivamente el significado de los resultados y deciden qué cuenta como éxito o fracaso.

Shadish et al. (1991) fueron más allá y sostuvieron que la evaluación había pasado de ser una práctica aplicada a consolidarse como una disciplina emergente, con teorías, modelos y fundamentos propios. En esta línea, Patton (2008, 2021) ha defendido la evaluación basada en el uso (*utilization-focused evaluation*), que pone en el centro a los usuarios de la información y orienta todo el proceso a que los resultados sean comprensibles, pertinentes y aplicables para la toma de decisiones.

1.1.2. La pluralidad metodológica

La diversidad de enfoques refleja la pluralidad de la investigación evaluativa. Pérez Juste (2000) destaca que la educación puede entenderse como arte, ciencia o tecnología, lo que justifica que la evaluación no deba restringirse a un único método. Esta perspectiva resalta la importancia de la complementariedad metodológica, en la que se integran técnicas cuantitativas, cualitativas, históricas y comparativas según la naturaleza del objeto y los objetivos de la evaluación.

Alvira (1991) identifica varios factores que han contribuido a conformar la investigación evaluativa como un campo específico:

- La creciente interrelación entre programación y evaluación, que hoy se desarrollan de manera paralela.
- La consideración de la evaluación como un proceso que va más allá de la investigación convencional, al tener como finalidad la toma de decisiones prácticas.
- La implicación de las partes interesadas, a través de la investigación-acción y la observación participante.
- La integración de estrategias diversas, cuantitativas y cualitativas, que permiten captar la complejidad de los programas.

Este carácter plural ha sido reforzado por el auge de los métodos mixtos, que combinan el rigor de la medición cuantitativa con la riqueza interpretativa de los enfoques cualitativos. Autores como Mertens (2019) y Creswell & Plano Clark (2018) han mostrado cómo la triangulación metodológica contribuye a ofrecer una visión más completa de los programas, a la vez que favorece la participación de los actores y la inclusión de perspectivas diversas.

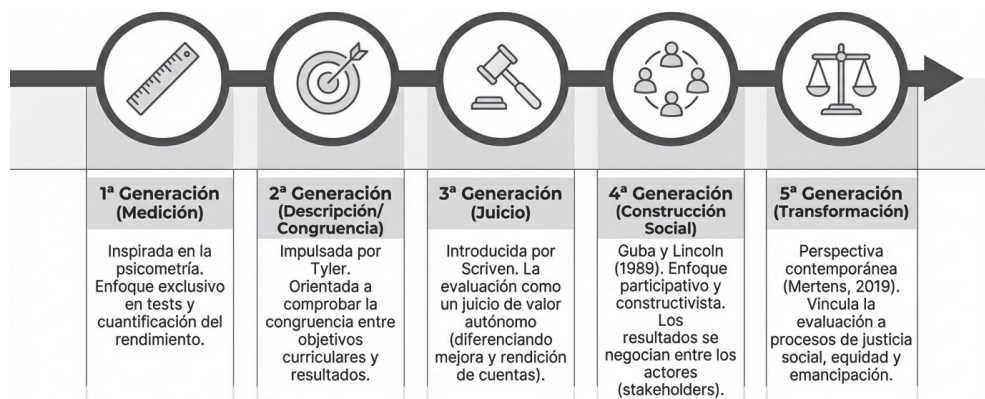
1.1.3. Generaciones de la evaluación: de la medición a la construcción social

Una de las formas más conocidas de sistematizar la evolución de la investigación evaluativa es la propuesta de Egon Guba y Yvonna Lincoln (1989), quienes identificaron distintas generaciones:

1. Primera generación: la evaluación como medición. Inspirada en la psicometría, centrada en los tests y en la cuantificación del rendimiento.
2. Segunda generación: la evaluación como congruencia entre objetivos y resultados, impulsada por Tyler, con un enfoque curricular.
3. Tercera generación: la evaluación como juicio de valor autónomo, defendida por Scriven, que introdujo la distinción entre evaluación formativa y sumativa.
4. Cuarta generación: la evaluación como construcción social, participativa y negociada, orientada a la pluralidad de intereses y significados.

Figura 1.1

Evolución de la disciplina



Autores posteriores han hablado incluso de una quinta generación, que vincula la evaluación a procesos de transformación social, justicia y equidad, incorporando perspectivas críticas y emancipadoras (Mertens, 2019).

Este recorrido muestra que la investigación evaluativa no es un campo estático, sino un ámbito en permanente evolución, sensible a los cambios sociales y a las demandas de transparencia, calidad y mejora continua en educación.

1.1.4. La investigación evaluativa como disciplina

Más allá de las clasificaciones históricas, lo cierto es que la investigación evaluativa ha desarrollado un cuerpo de conocimiento propio. Fitzpatrick et al. (2019) destacan que la evaluación se apoya en una terminología específica, en un conjunto de herramientas conceptuales y analíticas y en una comunidad de práctica profesional que comparte marcos, códigos éticos y estándares de calidad.

En este sentido, la evaluación no puede entenderse solo como un método, sino como una disciplina transdisciplinar. Scriven (1991) ya defendía que debía emanciparse de la investigación educativa o social, consolidándose como un campo autónomo, aplicable a cualquier área donde se implementen programas. Esto se refleja en el desarrollo de estándares internacionales, como las normas del *Joint Committee on Standards for Educational Evaluation*, que establecen criterios de utilidad, viabilidad/factibilidad, ética/legitimidad, precisión/rigor y rendición de cuentas.

1.1.5. La evaluación en el ámbito iberoamericano y español

Si bien los orígenes de la investigación evaluativa se encuentran en Estados Unidos, su desarrollo se extendió progresivamente a Europa y América Latina. En el caso de España, la consolidación de este campo se intensificó a partir de las décadas de 1980 y 1990, en paralelo al fortalecimiento del sistema educativo democrático y a la creciente necesidad de garantizar la calidad, la equidad y la rendición de cuentas en las políticas públicas.

Diversos autores han contribuido a situar la evaluación de programas en el centro de la reflexión pedagógica. Entre ellos destacan De la Orden (1985, 2009), Escudero (2003), Martínez-Mediano (2004, 2017, 2022, 2025) o Santos-Guerra (2010). En particular, De la Orden subrayó que la evaluación constituye un elemento inseparable de los procesos de planificación y mejora educativa, mientras que Escudero puso de relieve su dimensión política y social, al entenderla tanto como mecanismo de legitimación de las políticas educativas como herramienta para la transformación de la práctica pedagógica.

En esta misma línea, investigaciones aplicadas más recientes muestran la creciente importancia de implementar la investigación evaluativa en ámbitos socioeducativos concretos. El trabajo de Martínez-Mediano (2025), por ejemplo, analiza el uso de los modelos lógicos para el diseño y evaluación de programas educativos, destacando cómo estos permiten articular de forma sistemática el análisis de necesidades, la planificación de intervenciones, la implementación de acciones y la valoración de resultados e impactos. Desde esta perspectiva, la evaluación se concibe como un proceso continuo orientado a la mejora, que combina criterios de calidad, indicadores y metodologías mixtas para generar evidencias útiles para la toma de decisiones.

Asimismo, investigaciones empíricas como la de Almagro-Gavira (2017) evidencian la relevancia de aplicar la investigación evaluativa en contextos educativos y sociales específicos. Este tipo de estudios muestran que la evaluación puede funcionar no solo como herramienta de control o acreditación, sino también como un dispositivo de innovación y aprendizaje organizativo. En conjunto, estas aportaciones indican que la investigación evaluativa en España ha alcanzado un grado notable de madurez académica y profesional, consolidándose como un campo disciplinar propio estrechamente vinculado a la práctica educativa y a la mejora de las políticas públicas.

En el contexto latinoamericano, también se han producido contribuciones relevantes que han ampliado el alcance de la investigación evaluativa. Entre ellas destaca el trabajo de Casassus (2000), quien introdujo una concepción de la calidad educativa estrechamente vinculada a la equidad y al desarrollo de sistemas educativos inclusivos. Además, diversos países como México, Chile o Colombia han incorporado la evaluación de programas a gran escala como parte de sus políticas de innovación y reforma educativa, convirtiéndola en un componente clave para la toma de decisiones y la mejora de los sistemas educativos.

1.1.6. Retos actuales de la investigación evaluativa

La investigación evaluativa contemporánea se enfrenta a múltiples retos, entre ellos, la necesidad de equilibrar la función de rendición de cuentas con la de mejora. Mientras que las demandas sociales y políticas presionan hacia evaluaciones sumativas y de control, la práctica educativa requiere evaluaciones formativas, participativas y útiles para la innovación (Stufflebeam & Shinkfield, 2007).

Otro reto es la incorporación de tecnologías emergentes, como la inteligencia artificial, los sistemas de *big data* y la analítica de aprendizaje. Estos instrumentos ofrecen nuevas posibilidades para el seguimiento de programas, pero plantean dilemas éticos relacionados con la privacidad, la equidad y el sesgo algorítmico (Zhao et al., 2024).

Finalmente, la investigación evaluativa debe responder a la creciente diversidad cultural y social, adoptando enfoques interculturales y sensibles al contexto. Mertens (2019) insiste en que la evaluación no puede ser culturalmente neutra: debe reconocer la pluralidad de valores, voces y significados, especialmente en sociedades marcadas por la desigualdad y la exclusión.

En suma, la investigación evaluativa constituye un campo en expansión, que combina rigor metodológico, sensibilidad social y utilidad práctica. Surgida en un contexto de programas sociales masivos, ha evolucionado hasta convertirse en una disciplina transdisciplinar que integra diversos paradigmas, metodologías y enfoques. Dentro de este marco general, la evaluación de programas aparece como su expresión más característica, destinada a analizar la coherencia, el desarrollo y los resultados de intervenciones concretas, y su valor radica, no solo en ofrecer información fiable, sino en

contribuir a la mejora de los programas, al aprendizaje institucional y a la transformación social. En este sentido, la investigación evaluativa no es un lujo académico, sino una necesidad para garantizar la calidad, la pertinencia y la equidad de las políticas y prácticas educativas contemporáneas.

1.2. EL CONCEPTO DE EVALUACIÓN DE PROGRAMAS

La evaluación de programas constituye la expresión más concreta y aplicada de la investigación evaluativa en el ámbito educativo y social. Si la investigación evaluativa puede entenderse como un campo teórico y metodológico general, la evaluación de programas representa su instrumento de aplicación directa, orientado a analizar intervenciones específicas diseñadas para atender necesidades sociales o educativas. Esta dualidad entre marco teórico y práctica aplicada es fundamental para comprender su naturaleza compleja y multifacética.

1.2.1. Definición y evolución conceptual

La conceptualización de la evaluación de programas ha experimentado una notable evolución epistemológica. Una de las formulaciones más influyentes, coherente con los estándares del *Joint Committee on Standards for Educational Evaluation*, la define como el enjuiciamiento sistemático de la valía o el mérito de un objeto (Stufflebeam & Shinkfield, 2002). Esta definición nuclear, aparentemente sencilla, encapsula una profunda complejidad conceptual que merece ser desglosada.

El enjuiciamiento sistemático distingue a la evaluación profesional de la mera opinión o control administrativo informal. Implica un proceso riguroso, planificado y replicable de recolección, análisis e interpretación de datos, lo que confiere validez, confiabilidad y credibilidad a sus conclusiones. La sistematicidad constituye el antídoto contra la arbitrariedad y es lo que dota a la evaluación de su estatus de práctica científica.

La dimensión axiológica representa el corazón de la evaluación. Scriven (1991) realiza una contribución crucial al diferenciar analíticamente entre mérito –la calidad intrínseca del programa evaluada en función de sus propios objetivos y estándares de buena práctica– y valía –la utilidad, significatividad o importancia del programa en un contexto socioeconómico, cultural y político específico. Esta distinción es fundamental: un programa de formación de profesorado de Historia puede ser meritorio (produce docentes excelentes) pero carecer de valía si existe un excedente de tales profesionales en la sociedad.