

✓ Uma revisão dos indicadores de avaliação da formação pós-graduada em Espanha.

Javier Caballero Ferrándiz

Universidad Complutense de Madrid (UCM).

Resumo: Os indicadores são a espinha dorsal de uma avaliação. Tendo em conta os diferentes modelos de avaliação da formação pós-graduada em Espanha, no trabalho seguinte procede-se a uma revisão dos indicadores de avaliação segundo os seus elementos e aspectos técnicos. Para tanto, será realizada uma breve revisão teórica sobre algumas das diferentes propostas encontradas em torno da avaliação da formação pós-graduada e, de forma mais geral, sobre os métodos de avaliação com foco em processos e aqueles com foco em resultados. Posteriormente, procede-se a uma análise descritiva dos diferentes desenhos de avaliação da formação pós-graduada em Espanha: tanto a ANECA, como também os Rankings Universitários e o Mestrado em Avaliação (UCM), segundo os seus diferentes elementos que os compõem.

Palavras-chave: Avaliação educacional, Avaliação de programas, Indicadores de avaliação, Questões de avaliação, Métodos quantitativos.

EN A review of the Evaluation Indicators of Post-Graduate Training in Spain.

Abstract: The indicators are the backbone of an evaluation. Taking into account the different models of evaluation of postgraduate training in Spain, in the following work a review of the evaluation indicators is carried out according to their elements and technical aspects. For this, a brief theoretical review will be made of some of the different proposals found around the evaluation of postgraduate training, and more generally the evaluation methods focused on processes and those that focus on results. Subsequently, a descriptive analysis of the different evaluation designs of postgraduate training in Spain is carried out: both the ANECA, as well as the University Rankings and the Master of Evaluation (UCM), according to their different elements that comprise it.

Key-words: Educational evaluation, Program evaluation, Evaluation indicators, Evaluation questions, Quantitative methods.

ES Una revisión de los indicadores de evaluación de la formación de post-grado en España.

Resumen: Los indicadores son la osamenta de una evaluación. Teniendo en cuenta los distintos modelos de evaluación de la formación de postgrado en España, en el siguiente trabajo se realiza una revisión de los indicadores de evaluación de acuerdo con sus elementos y aspectos técnicos. Para ello se realizará un breve repaso teórico a algunas de las diferentes propuestas halladas en torno a la evaluación de la formación de postgrado, y de forma más general a los métodos de evaluación centrados en procesos y los que se focalizan en resultados. Posteriormente se realiza un análisis descriptivo de los distintos diseños de evaluación de la formación de postgrado en España: tanto la ANECA, como los Rankings Universitarios y el Máster de Evaluación (UCM), de acuerdo con sus diferentes elementos que la forman.

Palabras-clave: Evaluación educativa, Evaluación de programas, Indicadores de evaluación, Preguntas de evaluación, Métodos cuantitativos.

FR Une revue des indicateurs d'évaluation de la formation postuniversitaire en Espagne.

Résumé: Les indicateurs sont l'épine dorsale d'une évaluation. Compte tenu des différents modèles d'évaluation de la formation postuniversitaire en Espagne, dans le travail suivant, une revue des indicateurs d'évaluation est effectuée en fonction de leurs éléments et aspects techniques. Pour ce faire, une brève revue théorique sera menée sur quelques-unes des différentes propositions trouvées autour de l'évaluation de la formation post-graduée, et plus généralement sur les méthodes d'évaluation axées sur les processus et celles axées sur les résultats. Par la suite, une analyse descriptive des différents modèles d'évaluation de la formation postdoctorale en Espagne est réalisée : à la fois l'ANECA, ainsi que les classements universitaires et le Master of Evaluation (UCM), selon leurs différents éléments qui le composent.

Mots-clés: Évaluation pédagogique, Évaluation de programme, Indicateurs d'évaluation, Questions d'évaluation, Méthodes quantitatives.

1. INTRODUCCIÓN:

El presente artículo tratará de realizar una pequeña revisión de los indicadores de evaluación utilizados por tres modelos diferentes de evaluación de la formación de postgrado a partir de la metodología del Máster en Evaluación de Programas y Políticas Públicas de la Universidad Complutense de Madrid (UCM): la conocida como Rayuela.

Estos modelos, aplicados en concreto a las enseñanzas de Máster o Título Propio, serán: el de la ANECA –el generalmente más conocido y establecido en nuestro país como modelo de referencia de las enseñanzas universitarias–, el de Ránkings Universitarios –no tan conocido a nivel nacional, pero sí a nivel internacional–, y el del Máster de Evaluación (UCM) –como una apuesta propia para evaluar con criterios y medios propios el desempeño de esta titulación, así como para hacer extensivos sus métodos de concebir la evaluación de programas y políticas públicas–.

Como se comentará más en detalle en el cuarto apartado, la elección de estos tres modelos está motivada por la de comparar propuestas diferenciales según su enfoque de indicadores de resultados o de procesos. Cabe decir que este trabajo sirve de pretexto, además, para mejorar y ampliar las miras de la propia evaluación que realiza el Máster en Evaluación (UCM) a partir de su comparación con otros modelos análogos.

Pero hablando de la evaluación educativa, o la evaluación de la práctica de la enseñanza, es importante señalar la diferencia que se hace en el habla inglesa entre “assessment” –evaluación del alumnado mediante pruebas, test, controles, etc.– y “evaluation” –evaluación de programas y políticas públicas–, teniendo en cuenta que en este trabajo siempre se hará referencia más a la segunda que a la primera. En este sentido, es importante recalcar el significado de esa frase que se viene repitiendo muchas veces en el Máster de Evaluación: “evaluamos programas, no personas”.

El objeto, por tanto, será evaluar los indicadores de los propios programas de evaluación de la formación de postgrado desde el punto de vista de cómo están concebidos teóricamente y nunca los procedimientos que se utilizan para medir o evaluar –en el sentido inglés de “assessment”– el desempeño o los esfuerzos del alumnado por aprobar una asignatura en el campo de la evaluación del aprendizaje educativo.

Para ello se realizará una breve presentación de la metodología utilizada para la evaluación, a través de la llamada Rayuela de evaluación, con cuño propio del Máster en Evaluación de Programas y Políticas Públicas (UCM). De forma breve además se tratará de presentar el marco teórico de la evaluación de la formación de postgrado y los tres modelos de evaluación comentados. Con ello se irá comparando a lo largo de todo este proceso los diseños y presentando, al final de cada apartado, un breve cuadro resumen donde se puedan ver rápidamente las diferencias entre estos modelos.

Al final, en la discusión se aportará una breve reflexión acerca de cómo la aproximación metodológica de la evaluación condiciona el tipo de indicadores, estableciendo una senda típica y tópica entre la orientada hacia resultados y la orientada a la comprensión del programa. Y en ese sentido, se puede interpretar que no es que sean modelos de evaluación “cortos de miras”, es el propósito y la capacidad de acción de sus sujetos los que la marcan.

2. MARCO TEÓRICO:

Resulta difícil encontrar referencias tan concretas sobre la evaluación de la formación de postgrado, salvo como las que se encuentran en el Handbook de SAGE sobre evaluación educativa en la educación superior (Secolsky & Denison, 2012) o algunas como la evaluación de un máster online (Sancer, Baturay & Fadde, 2013). Pero quizás la más pertinente, dado el caso, sea la que aquí reseñaremos de David Nevo en el Handbook de SAGE sobre evaluación de políticas, programas y prácticas (Shaw, Greene & Mark, 2006).

Este autor anglosajón distingue entre diversos objetos que son de evaluación en la educación. En concreto enumera los siguientes: alumnado, profesorado, materiales de enseñanza, proyectos y programas educacionales y escuelas. Pero es el último el que más se asimila al objeto de indicadores teniendo en cuenta todos los siguientes elementos a evaluar (Nevo, 2006:445-457):

- La comunidad y el alumnado de la escuela: elementos contextuales como el nivel socioeconómico del alumnado y del barrio donde esta se encuentra, la visión de la escuela, etc.
- El personal de la escuela: profesorado, directoría, etc.
- Los recursos materiales: desde la pizarra y la tiza pasando por

los ordenadores, hasta el presupuesto, el espacio y todos los enseres, aunque también la limpieza y el nivel de ruido.

- Los programas educativos y las actividades: los programas principales y tanto aquellos dirigidos al alumnado destacado como con dificultades, las actividades extracurriculares, las disposiciones hechas para estudiantes con necesidades especiales, los proyectos que sean innovativos iniciados por la propia escuela u otras, así como aquellos proyectos que afecten a la comunidad o a los progenitores.
- Los logros de la escuela: el reclamo que tenga esta escuela, la satisfacción de los progenitores, los juicios de los inspectores de escuela, logros y premios de sus graduados, o premios ganados por la propia escuela.

Así pues, si los inicios de la evaluación en un primer momento se remontan a la definición inglesa de “assessment” –pasando por la evaluación educativa de Ralph Tyler– los desarrollos que actualmente entendemos como evaluación de programas y políticas públicas –así como de organizaciones, servicios públicos o incluso centros educativos– no se definieron como tal hasta el Modelo CIPP de Daniel Stufflebeam (Nevo, 2006: 441) al intentar incluir, de forma sistémica y sistemática, todos aquellos elementos que incluyen los insumos (estructura y contexto), los procesos y los resultados (aunque también productos).

Esta consideración sigue abriendo el foco actualmente incluso de la evaluación de los programas educacionales y las escuelas a elementos que van más allá de los resultados de informes como PISA (Programa para la Evaluación Internacional de Alumnos) –y que se quedarían en el nivel de evaluación del “assessment”–. Como comenta el autor, en Gran Bretaña no solo se tienen en cuenta los resultados de los test, también se evalúan los métodos de enseñanza del profesorado, el currículum, o incluso elementos como el clima escolar (Nevo, 2006: 454), demostrando la pertinencia de incluir aspectos no solo de resultados.

Por citar solo un caso concreto, con el que continuar con el ejemplo, a continuación se hace una pequeña referencia a un cuadro aplicado a la evaluación de una escuela e inspirado en el modelo CIPP que puede servir de consideración para tener en cuenta todos aquellos elementos de contexto, estructura, procesos y productos/resultados que, al igual que lo haría en una escuela, son aplicables a la evaluación

de una formación de postgrado (Aziz, Mahmood & Rehman, 2018).

Sin embargo, dada la sistematicidad de este sistema, es más bien frecuente ver evaluaciones e indicadores que se centran solo en aspectos de resultados –sobre todo en materia de políticas públicas más que de programas–. Es importante ahora preguntarse por qué se produce esta cuestión y por qué no se tienen en cuenta las virtudes que un modelo como este son capaces de dar. ¿Por qué no abrir la caja negra y comprender los mecanismos que generan los resultados? (Ligero, 2017: 36).

Como veremos en los siguientes apartados a través de la comparación de los tres modelos, esta es una cuestión fundamental a la hora de comprender que existe una cierta relación entre los indicadores evaluación y sus propósitos –entre sus medios y sus fines–. Y que es necesaria de comprender por el mero hecho por el que se plantea: es así como por ejemplo David Nevo establece una clara relación entre una evaluación interna-formativa y una externa-sumativa. Este autor comenta que “por un lado, las escuelas pueden estar interesadas en una visión global de sus actividades educacionales y administrativas para mejorar su funcionamiento y ejecución. Por el otro, una demanda puede requerir una demostración del mérito de la escuela y el grado en que consigue sus resultados y la satisface las necesidades de sus clientes”, estableciendo la diferenciación entre una evaluación interna y una externa, y también entre formativa y sumativa, a partir de los propósitos y el contexto en el que se formulan (Nevo, 2006: 454).

Sobre esta cuestión incidirán de nuevo todos aquellos elementos contextuales que no tienen que ver solo con el programa, sino con el requerimiento de la demanda, aquellas personas que participan en la evaluación, o incluso el contexto social y económico que influye en la formación de postgrado y que afecta a su posible evaluación. La relación entre fines y medios no es solo una cuestión de método, es ante todo una cuestión de contexto: una cuestión política. Y justamente afecta también al planteamiento de sus indicadores.

3. METODOLOGÍA:

Siguiendo la guía práctica de evaluación propuesta por el “*Título Propio en Evaluación de Programas y Políticas Públicas*” de la Universidad Complutense de Madrid (UCM) –más conocida como la “Rayuela”, o en su versión en inglés como *hopscoth*, por su

referencia al juego infantil de las casillas– se realiza un comentario crítico en torno a los indicadores que comprenden el diseño de las evaluaciones de la formación de postgrado como un ejercicio de reflexión y comprobación para hacer una evaluación consciente (Ligero, Porta, Muñoz, y Bustelo, 2019: 1)

Inicialmente esta herramienta se concibe en 12 fases que van desde las primeras preguntas de “¿para qué se evalúa?” y “¿para quién?”, pasando por la acotación del objeto de evaluación, la elección de metodológica, la definición de los indicadores, la definición de las fuentes de información, el análisis de los datos, entre otras; hasta la revisión de los indicadores de evaluación de cada programa.

De forma práctica la Rayuela se nos ofrece como un checklist, como un ejercicio de comprobación, a través de distintas preguntas que resultan muy agudas en torno a la práctica de la evaluación. Pero también como un ejercicio de reflexión con lo que ello supone: pararse en cada una de las preguntas y pensar una respuesta que vaya más allá de la mera comprobación.

Cada una de las fases son preguntas, por tanto, siempre sugerentes y abiertas a más de una respuesta –o por lo menos a una respuesta larga, no una de “sí” o “no”–. Por ello “el proceso de evaluación se convierte en un ejercicio consciente de la diversidad de opciones, productos y usos [...] el producto final es un diseño de evaluación sensible, y por lo tanto, específico y conciso” (Ligero, Porta, Muñoz, y Bustelo, 2019: 7).

Comentar solo como apunte que la idea de esta herramienta nace justamente de un esquema que se utilizaba para las clases del Máster con la intención de ayudar a “diseñar, repensar y revistar la implementación de los procesos de evaluativos” (Ligero, Porta, Muñoz, y Bustelo, 2019: 4) así como de ordenar el propio proceso de aprendizaje de la evaluación por parte del alumnado. Es por tanto una herramienta intrínsecamente pedagógica y dispuesta a su uso.

4. ANÁLISIS:

En un primer momento, se trató de revisar el modelo de evaluación desarrollado por el Máster de Evaluación –que no tiene carácter de Máster Oficial como tal, sino de Título Propio– con el que regula el resto de las enseñanzas de Máster Oficial –el de la ANECA– ya que el

segundo es el que rige todas las enseñanzas de postgrado oficial en España. Y ese fue el criterio inicialmente establecido, aunque como se ha comentado anteriormente, este planteaba la necesidad de contar con algún otro diseño adicional con el que enriquecerse.

Finalmente se eligieron los Ránkings Universitarios teniendo en cuenta las virtudes derivadas de su carácter completamente diferenciador en comparación con la ANECA y el Máster en Evaluación (UCM): una herramienta de carácter internacional principalmente orientada a comparar los resultados de las universidades en términos de producción académica e internacionalización de sus productos y personas.

4.1. ANECA (ACREDITA):

Con la implantación del sistema Bolonia para el desarrollo de un Espacio Europeo común de Educación Superior, se produjeron distintos cambios estructurales que afectaron al sistema educativo europeo y a las universidades españolas.

Este es el propósito bajo el cual funciona la Agencia Nacional de Evaluación de la Calidad y Acreditación –más comúnmente conocida como ANECA–, creada en 2002 como un organismo homólogo a la European Association for Quality Assurance in Higher Education (ENQA), y que tiene como objetivo dotar al sistema español de un proceso que trate de garantizar su calidad y mejora continua a través de la certificación y acreditación de las enseñanzas universitarias.

La ANECA cuenta con distintos programas que tratan de cubrir el amplio proceso de que culmina en una evaluación, pasando desde la evaluación de diseño (VERIFICA), por el monitoreo y el seguimiento (MONITOR), hasta la evaluación de resultados y la rendición de cuentas (ACREDITA). Aunque también, son conocidos los programas de evaluación de las instituciones académicas (AUDIT), y sobre todo el programa de evaluación del profesorado (DOCENTIA).

Todos estos programas nacen con la voluntad de garantizar la calidad en la formación del alumnado de las universidades a través de la mejora continua de los procesos y los recursos que las forman. Sin embargo, es el de ACREDITA el que tiene interés en este artículo sobre indicadores dada su aplicación a la evaluación de las enseñanzas de postgrado con carácter tanto de su implementación como de resultados.

El programa ACREDITA –quizás el principal ya que es el que se enfoca hacia la rendición de cuentas– es el que regula el proceso de acreditación de los Grados, Másteres y Doctorados en función del grado de cumplimiento de los criterios y estándares establecidos. Estos criterios se dividen entre aquellos de la dimensión de gestión del título, los recursos y los resultados:

Cada una de estas dimensiones se subdivide en 7 criterios, de forma que la primera engloba también la organización y su desarrollo, la información y transparencia, y el sistema de garantía interno de calidad. La segunda dimensión incluye el personal académico y el personal de apoyo, así como los recursos materiales y servicios. Por último, la dimensión de resultados abarca tanto estos resultados de aprendizaje como otros indicadores.

Teniendo en cuenta los criterios, el resultado es una matriz de indicadores formada para la evaluación de ACREDITA con 24 indicadores. Sin embargo, teniendo en cuenta que los aspectos a valorar pueden ser de 2-3 de media, en realidad se están respondiendo a entre 50 y 75 indicadores. La matriz de autoevaluación ACREDITA cuenta en total con 68 indicadores, de los cuales éstos se pueden dividir en tres subindicadores:

- Información en la que se basa el análisis (fuentes).
- Aspectos a valorar para el cumplimiento de la directriz (mejora).
 - Valoración semicuantitativa: se supera excelentemente, se alcanza, se alcanza parcialmente o no se alcanza.
 - Valoración cualitativa.
- Preguntas a modo de reflexión (exhaustivo checklist).

El primer apartado es un requerimiento sobre las fuentes de información en las que se basa la respuesta de cada indicador: estos pueden ser informes, encuestas, entrevistas u otros documentos. Como se puede ver, el segundo supone una autovaloración en términos ordinales de cada uno de los ítems, así como una valoración cualitativa. El último apartado incluye un largo conjunto de preguntas a modo de verificación para comprobar el grado de consecución de los procesos. Suele ser un listado bastante exhaustivo, llegando hasta las 20 preguntas para conformar un único indicador.

Sin embargo, el informe de autoevaluación es solo uno de los otros 11 documentos que forman el dossier de evaluación. Por ese mismo

motivo la ANECA no ha estado exenta de múltiples críticas y objeciones desde su nacimiento, en relación con su falta de transparencia, exceso de burocracia, la reducción de plantillas y presupuestos, en detrimento de las ciencias no experimentales, o la falta de objetividad y distanciamiento (VV. AA., 2013).

El proceso de evaluación por el cual se procede a la renovación de la acreditación de los títulos es el de una primera autoevaluación donde la universidad describe y valora la situación del título respecto a los criterios establecidos, y una segunda evaluación externa por parte de un panel de expertos, que analizan el autoinforme y realizan una visita al centro, y por último una evaluación por parte de la comisión de acreditación de la ANECA a modo de juicio final sumativo.

Pero es que además esta se complementa con un proceso por el cual participan personal externo a la propia universidad con la intención de realizar una visita in situ sobre los niveles de cumplimiento de lo establecido en las memorias de verificación del título. Este participa en tres momentos clave.

El proceso incluye varias fases. Antes de la visita, recepción y estudio individual del informe de autoevaluación, reuniones con los miembros del panel y selección de las asignaturas, y por último una reunión preparatoria a la visita. Esta reúne varias fases relacionadas con el desarrollo de la visita, entrevistas con diferentes colectivos como puede ser el alumnado, revisión documental e incluso visita de las instalaciones. Después de la visita es básicamente la elaboración del informe del panel de expertos.

4.2. Ránkings Universitarios (QS y THE):

Los generalmente conocidos como Ránkings Universitarios ordenan a las principales universidades e instituciones de educación superior de acuerdo con un posicionamiento bibliométrico de acuerdo con el impacto de sus publicaciones académicas y su impacto económico en la sociedad. El objetivo de estas listas es dar a conocer públicamente la calidad, en términos de resultados, de estas universidades. Principalmente se encuentran dos tipos: QS y THE.

Por tratar de resumir brevemente, QS es la original y es a la que muchas veces sigue siendo referencia puesto que no ha cambiado en su metodología de tipo bibliométrico desde 2004. Sin embargo, THE

parte de esta primera para actualizarse con contenidos más allá de los bibliométricos a partir de 2009.

Es esta última la que resulta más interesante para aplicar a la evaluación de formaciones de postgrado al tener un carácter mucho más amplio y también de los procesos educativos. Es decir, y con ánimo de recalcarlo de nuevo: aunque la QS sigue siendo la más generalizada, la THE permite realizar una evaluación que no tenga en cuenta solo la producción científica e investigadora en forma de citas, sino también los procesos de enseñanza.

QS World University Rankings

El Ránking Universitario Global QS sirve de referencia para el top de universidades en materias específicas. A continuación, se resumen brevemente los 6 indicadores y sus respectivos pesos que estructuran esta metodología de posicionamiento de los principales centros de formación de postgrado:

El indicador de la reputación laboral y académica se basa en un ránking que el propio personal académico realiza entorno a una lista de 10 nacionales y 30 internacionales instituciones que consideran adecuadas para que estudiantes y ellos mismos estudien o trabajen, y las disciplinas a las que les gustaría participar. El resultado de estos cruces da como resultado la excelencia de cada universidad.

Los siguientes indicadores incluyen además estos elementos de acuerdo con los objetivos con los que se plantean:

- Reputación académica (40%).
- Citas por facultad (20%): según la gran base de datos de Elviseo y en un período de 5 años vista.
- Ratio de profesorado y alumnado por facultad (20%): las ratios son un criterio ajustado, del cual conocemos el umbral aunque este puede que sea 1 a 1, según el cual la medida no puede ser ni por exceso ni por defecto.
- Reputación inserción laboral (10%): el alumnado concibe la universidad como un medio para acceder al mercado laboral.
- Ratio de internacionalización del profesorado (5%): favoreciendo el intercambio de buenas prácticas entre el profesorado.
- Ratio de internacionalización del alumnado (5%): capacidad de atracción de público internacional y formación de marca,

y competencias como la conciencia global y la simpatía internacional para una mayor empleabilidad del alumnado.

El famoso índice-H –denominado así en mención a su propio autor, el físico argentino Jorge E. Hirsch– se basa en el número de artículos más citados y el número de citas que recibe un artículo por parte de otras personas en otros artículos. Como referencia sirve la base de datos de publicaciones Scopus en un período de cinco años hacia atrás.

THE World University Rankings

Sin embargo, se puede decir que la guía de ránking más utilizada en España es la británica THE publicada por el diario *Times*, al que se le añaden algunos criterios no bibliométricos como: el número de estudiantes matriculados, la proporción entre estudiantes y graduados, la proporción de personal académico doctorado, el número de cursos impartidos, etc. Se consideran todos estos como indicadores de producto importantes en tanto en cuanto hacen referencia a elementos de proceso como son la propia enseñanza y la calidad de la investigación.

El Ránking de las Universidades del Mundo Times Educación Superior (THE) trata de juzgar la actividad investigadora realizada por las universidades en sus materias principales: enseñanza, investigación, citas, transferencia del conocimiento y perspectiva internacional. Este trata de establecer comparaciones más comprensivas y ajustadas a partir de 13 indicadores calibrados por estudiantes, personal académico, staff de las universidades, industrias y gobiernos. Evaluación basa en criterios, en concreto 5 indicadores que son los siguientes:

La enseñanza comprende de nuevo la propia encuesta recíproca sobre reputación, el ratio de profesorado respecto al alumnado, el ratio de doctorandos respecto a graduados, otra ratio sobre doctorados con mención respecto al profesorado académico, y un último indicador sobre los insumos institucionales. La encuesta pregunta por el prestigio de las universidades percibido a través de su enseñanza en la medida en que los graduados pueden argumentar que esta será efectiva y útil para otros futuros graduados. Quedando bastante claros los cuatro

primeros indicadores, el último queda controlado a través de la paridad del poder adquisitivo (PPA) que es una de las medidas más adecuadas para comparar la producción de bienes y servicios, puesto que toma en cuenta las variaciones de precios. Este indicador elimina la ilusión monetaria ligada a la variación de los tipos de cambio.

Sobre la investigación este indicador se encuentra una nueva encuesta de reputación investigadora, los insumos de investigación con que cuenta cada facultad, y la productividad investigadora según la base de datos académica escalada para el tamaño institucional y normalizada para tema. Este indicador resulta un poco complicado para investigaciones que no tienen una transformación directa en términos económicos y monetarios.

Las citas hacen referencia a lo que se produce y cómo se reproduce entre la comunidad académica. Controlada para que las materias con mayor volumen de citación, como las de ciencias naturales, no tengan más peso respecto a las demás. Contribuyen a la expansión del conocimiento científico y al intercambio de buenas prácticas entre la comunidad investigadora.

La perspectiva internacional asume la ratio de alumnado internacional sobre el nacional, la ratio de persona académico internacional sobre el nacional, y la colaboración que se establece de tipo internacional. En estos aspectos se señala la capacidad de una universidad de atraer público de otros lugares y la producción que tiene coautoría de tipo internacional.

La transferencia de conocimiento a la industria queda definida por la capacidad de una universidad para ayudar a las empresas con innovaciones, inventos y avances. También por la proporción entre el número de incomes que realiza respecto a los profesores que tiene partido por el (PPA). Solo se aplica a entornos académicos donde se pueden valorar en términos monetarios.

4.3. Máster evaluación (UCM).

El Máster Propio en Evaluación de Programas y Políticas Públicas de la Universidad Complutense de Madrid (UCM) se imparte ininterrumpidamente desde el año 2002. Es un programa teórico-práctico dirigido a la adquisición de conocimientos y competencias para responder las necesidades y requerimientos de evaluación

del personal directivo y técnico de las administraciones públicas, privadas y del tercer sector; está pensado también para profesionales de la consultoría y para académicos que quieran especializarse en estas áreas de estudio.

El Máster está posicionado como una formación completa, actualizada y de reconocido prestigio, pudiendo garantizar tras la experiencia exitosa de 17 promociones y de más de 400 personas formadas, que las personas que los cursan adquieren las competencias y conocimientos necesarios para diseñar, dirigir, realizar y encargar evaluaciones, así como analizar críticamente las mismas.

El modelo de evaluación desarrollado por el Máster en Evaluación (UCM) comprende una serie de técnicas, tanto cualitativas como cuantitativas, que abarcan desde algunas entrevistas a personas que se consideran clave en el funcionamiento del mismo, como un cuestionario dirigido a distintos públicos que participan tanto en el lado de la docencia como en el del aprendizaje. Se ordenan en tres fases simultáneas con la intención de que las técnicas se vayan complementando.

El modelo lógico del programa plantea indicadores a nivel de estructura, procesos, productos y resultados con la misma intencionalidad que otros modelos sistémicos como el de Stufflebeam, que reúne elementos basando tanto en la teoría del cambio (Weiss, 1998) y la evaluación a la cooperación al desarrollo, como mucha pedagogía y saber hacer (MAE, 2007). A continuación, se añade un pequeño extracto de la matriz de evaluación:

5. DISCUSIÓN:

En este apartado se aborda una de las cuestiones principales: los indicadores. O más bien, el proceso por el cual se construyen aquellas preguntas, criterios, estándares y sistemas de medida que son el centro de la evaluación generalmente planteados en la matriz de evaluación. Esta cuestión es importante ya que lo que no está en la matriz, no existe.

Como comúnmente se nos dice, la evaluación consiste en hacer buenas preguntas y responderlas bien (Ligero, Porta, Muñoz, y Bustelo, 2019: 42). En principio la segunda cuestión puede ser bastante clara: se trata de evitar errores estadísticos como “falsos

positivos” o “falsos negativos” –los conocidos en estadística como error tipo I y tipo II–.

Sin embargo, y para que la segunda cuestión quede clara, hay que hacer referencia –por analogía estadística– a los errores tipo III: aquellos en los que el error está derivado de dar una buena respuesta a un problema mal planteado (Mitroff & Silvers, 2010). Por muy bien que sean respondidas las preguntas, si no son preguntas relevantes y planteadas de acuerdo con lo que verdaderamente se quiere conocer, estaremos cometiendo un error también. Hacer una pregunta ya forma parte de la respuesta.

Para realizar una buena operacionalización quizás sea de nuevo importante ser consciente de quiénes han planteado el problema a evaluar, quién ha diseñado este sistema para responder a las preguntas, o incluso diferenciar las distintas preguntas según los stakeholders que las propusieron (Red CIMAS, 2015). Esta cuestión permitirá tener en cuenta preguntas de acuerdo con los criterios de coste, influencia, incertidumbre y abundancia de información (Ligero, 2017: 79).

Haciendo referencia a cada uno de los modelos de la evaluación de la formación, sobre la que realiza la ANECA no se piensa que se haya hecho esta selección de los indicadores en términos de relevancia, coste, influencia, incertidumbre o abundancia de información. El resultado es una serie de preguntas complementarias a los indicadores que resultan demasiado exhaustivas. Tampoco se tiene constancia de qué actores han planteado cada pregunta. Resulta un diseño de evaluación poco transparente.

Sobre los Ránkings Universitarios, se comentan que los THE han contado con la participación de algunos stakeholders que han influido en la inclusión de algunos indicadores y sistemas de valoración. Aunque este proceso no quede del todo explícito, es importante señalar que toda clase de estos procesos se considera como relevantes y pertinentes. Sin embargo, se le da mucho peso a las valoraciones que hacen otros centros sobre la reputación académica de las mismas, no siendo este un indicador objetivo. En los ránkings THE este tiene menos peso, pasa del 40% a solo el 15%. Este indicador, al igual que el siguiente, tienen la misma clase de sesgo que se produce en las votaciones de Euro Visión: los países que forman clústers de afiliación –países de Europa del Este o países nórdicos-bálticos, a diferencia de otros como Reino Unido, Francia o España– se votan entre sí siguiendo el precepto de que “el amigo

de mi amigo es mi amigo, y el enemigo de mi amigo es mi amigo”. La lógica es que Reino Unido no va a votar a Alemania, o a Francia, pero Lituania votará a Eslovenia de la misma manera que Eslovaquia votará a Eslovenia y esta a Lituania. Y así alguno de estos países del clúster será el ganador.

Sobre la Evaluación del Mágister (UCM) se comenta que “la dificultad estuvo en pasar de la fase divergente –recopilación de todas las posibles necesidades informativas– a la fase convergente –selección de las necesidades informativas–. La aplicación de criterios como coste, influencia y probable abundancia de información propuestos por Cronbach fueron de mucha ayuda” (Rodríguez, 2019: 23). Así fue como la matriz se fue desbrozando, evitando preguntar cuestiones de poco valor y que engrosaran demasiado el trabajo de la evaluación, hasta llegar a solo 65 preguntas de evaluación que fueran pertinentes a la vez de relevantes.

5.1. ¿Las preguntas implican una valoración y no una mera indagación?

En este momento se hace patente la diferencia entre una mera indagación y aquellas preguntas que supongan una valoración a cerca del programa. Esto hace referencia a todas las preguntas incluían algún tipo de valoración del tipo “adecuada, suficiente, clara, de calidad, etc.” que permitiera distinguir a estas preguntas de evaluación de unas preguntas de investigación.

Las preguntas de la evaluación de la ANECA incluyen adecuadamente elementos de valoración como aquellos que tienen que ver con “adecuado”, “pertinente”, “suficiente”, “coherente”, “congruente”, “actualizado”, “implantación de acuerdo con”, “formalmente establecido”, etc. Aunque se encuentran también términos de forma conjunta, como “suficiente y adecuado”, que pueden confundir. En la medida de lo posible la pregunta debe de restringirse a un único elemento valorativo, así como un único aspecto a valorar evitando preguntas como: “¿el personal de apoyo, recursos materiales y servicios puestos a disposición del desarrollo del título son los adecuados en función de la naturaleza, modalidad de enseñanza-aprendizaje, número de estudiantes matriculados y competencias a adquirir por los mismos?”.

Los rankings no están planteados como una pregunta de evaluación, sino como meros receptáculos donde incluir datos y ratios respecto a cada uno de los ítems formulados. Esta cuestión puede dificultar el paso de los datos a una interpretación que sea capaz de realizar un juicio sobre el programa de acuerdo con el marco de la evaluación. Son preguntas incluso, como veremos en el próximo apartado, que pueden no hacer referencia al programa.

De nuevo en la evaluación del Máster de Evaluación (UCM) se hace una diferencia clara entre preguntas valorativas y preguntas de indagación, siendo las primeras definidas con términos como “suficiente”, “efectiva”, “correcta”, “equilibrada”, “adecuada”, “eficiente”, “a tiempo”, “de calidad”, o “completa”; a diferencia de otras preguntas abiertas o de indagación como la siguiente: “¿se ha generado algún tipo de mejora difícilmente medible en el desempeño profesional posterior?”. Esta cuestión evidencia que las preguntas de indagación pueden ser un complemento a las de valoración siempre y cuando se coloquen en el marco adecuado, como por ejemplo este de los resultados no previstos por la formación (free goal).

5.2. ¿Las preguntas están dirigidas al programa y no a sucesos o dinámicas sociales?

Es importante tener claros qué elementos forman parte del programa y cuáles no, del mismo modo que es importante no preguntar por cuestiones que son ajenas al programa. Es decir, sucesos sociales y estructurales que escapan al mismo. Las dinámicas sociales requieren de un estudio propio ya que, generalmente, hacen referencia a elementos que en realidad están fuera de la influencia del programa. Son lo que se llaman factores externos o variables estructurales.

Quizás nos atrevamos a decir que, el hecho de plantear preguntas que hagan referencia únicamente al programa es un criterio para hablar del impacto de la intervención, esto es: en qué medida los resultados son atribuibles al programa y no a otras dinámicas sociales (Mayne, 2011).

De cara a la interpretación, el plantear preguntas dirigidas al propio programa nos va a garantizar cierto sentido de causalidad, en la medida en que estamos seguras de que los resultados se atribuyen

directamente con los procesos –a partir de las preguntas– que se han formulado.

De forma acertada, la evaluación de la ANECA no incluye preguntas externas al programa. Todas ellas hacen a elementos inherentes a la formación como son su organización y desarrollo, los sistemas de información de transparencia y sistemas de calidad, el personal académico, el personal de apoyo, los recursos materiales y servicios, así como los propios resultados de aprendizaje y los indicadores de resultados. Quizás, y en todo caso, peca de no incluir elementos previos y contextuales al programa que, sin embargo, le están afectando tanto a sus procesos como a sus resultados.

De nuevo en los Rankings Universitarios esta cuestión resulta un poco complicada, pues las preguntas hacen referencia a productos que bien pueden estar relacionados con el programa o no: la internacionalización del profesorado o su mayor capacidad investigadora también puede estar relacionada con dinámicas sociales y económicas que van más allá de la institución universitaria, así como con relación a su transferencia conocimiento a la industria

Sobre el Máster de Evaluación (UCM), esta cuestión resultó bastante sencilla al hacer referencia todas las preguntas al propio programa. Aunque hay elementos relacionados con el origen el público que cursa la formación, se plantea una pregunta en relación con si este motivo fue uno por los que se eligió la misma enseñanza. Es decir, como una manera de plantear ese elemento aparentemente externo dentro de la estructura interna del programa.

5.3. ¿Las preguntas se diferencian de los indicadores?

Quizás el motivo más aparente por el cual preguntas e indicadores deben diferenciarse está relacionado con el hecho de que las preguntas de evaluación –en la manera que se formulan, con algún adjetivo valorativo tipo “adecuado” o “suficiente”– son ya un puente entre el indicador de facto –un hecho– y el juicio de la evaluación –una interpretación–. Por ello es tan importante plantear preguntas que ayuden a realizar ese tipo de transición hacia la interpretación desde los meros datos.

La gran diferencia es que, en cierta medida, la pregunta incluye el estándar con el que se valorará la respuesta. Y este estándar,

básicamente, se puede considerar como la suficiencia del criterio. Recordemos que al utilizar un criterio de valor hacemos realidad la premisa de que “evaluar es algo más que investigar, puesto que debe de emitirse un juicio de valor sobre el objeto que se está evaluando” (Alvira, 1991: 7). Un indicador básico será el de suficiencia, y su estándar el punto medio pues: en el punto medio está la suficiencia.

Pero sobre esta cuestión puede ser de utilidad la diferencia entre indicadores descriptivos y analíticos: “por indicador descriptivo se entiende aquel que trata de explicar o poner de manifiesto la posible regularidad existente en un conjunto de datos. Por indicador analítico se entiende aquel que trata de explicar, más allá de las regularidades, el valor de los datos en el análisis que se realiza” (González, 1986: 236).

En este sentido es importante avanzar hacia indicadores analíticos que “se refieren a las combinaciones que se pueden realizar con indicadores individuales (p.e., los ingresos medios familiares que se obtienen de la razón entre la suma de los ingresos de cada miembro de la familia y el número total de los mismos de la misma)” (Lundberg, 1949: citado en González, 1986: 238). Éstos generalmente planteados en forma de ratio.

Sobre la evaluación de la ANECA, este aspecto se cumple ya que cada pregunta cuenta con un estándar que sirve como punto de referencia para valorar el grado de cumplimiento del título evaluado. Este estándar se corresponde con la valoración semicuantitativa de suficiencia que ayuda hacia la interpretación del programa: se supera excelentemente, se alcanza, se alcanza parcialmente o no se alcanza.

Sobre los Ránkings Universitarios, y al no tratarse de una evaluación en su sentido estricto, no encontramos este tipo estándares ni ningún otro. Pero se ofrecen ratios al considerarse como indicadores analíticos que pueden ayudar a realizar el paso a la interpretación en mejor medida que los de tipo meramente descriptivo. Para empezar este resultado va a ser sobre 1, siendo el mínimo 0 y el máximo este 1. Estos ayudan, en cierto sentido, a la interpretación, aunque no se consideran como suficientes para la evaluación.

La Evaluación del Máster (UCM) plantea adecuadamente en las preguntas términos valorativos que se puedan considerar, como un indicador cualitativo, acerca de la adecuación, la pertinencia, la suficiencia, o la calidad de cada uno de los indicadores formulados. También hay algunas ratios considerables en relación con el uso de las redes sociales. Se observa, por tanto, una relación

complementaria entre indicadores y técnicas cualitativas y cuantitativas respectivamente.

5.4. ¿Las preguntas son sencillas?

Es importante señalarla que de cara a realizar una evaluación ni los neologismos ni las frases subordinadas pueden ayudar: las preguntas de evaluación deben de ser claras, sencillas, y a ser posibles de respuesta rápida y breve –sí o no–. Así de sencillo, valga la redundancia.

Pero que sea así no supone que por ello sea menos “científico” o menos riguroso. Justamente la sencillez es el método como ha tratado de avanzar la ciencia: descomponiendo lo complejo en elementos más sencillos (Wagensberg, 2018) Si una pregunta de evaluación no es capaz de descomponer todos aquellos elementos del programa en algo más sencillo difícilmente la respuesta podrá ser clara y precisa.

Aunque estas preguntas no solo deben ser sencillas para quien diseña la evaluación, también para quién luego tiene que responderlas en un cuestionario o el público en general. Hay que retomar la antigua diferencia comentada en antropología entre el lenguaje etic –el de quien investiga– y el lenguaje emic –el de la población a la que se dirige. La claridad de las preguntas está en el oído de quién las escucha.

También si son planteadas, entendiendo de una forma muy laxa y muy poco exigente –a diferencia de, como muy acertadamente, se realiza en otros campos como la psicología– los términos de validez y fiabilidad. En este sentido, se entendería la primera como la capacidad de traducir en preguntas de cuestionario aquello que realmente se quiere conocer –y no otra cosa o esta misma relacionada con otra–, y en términos de fiabilidad si las personas entienden lo mismo que nosotros entendemos al formular la pregunta y sus categorías de respuesta.

La evaluación realizada por la ANECA, aunque efectivamente resulta muy exhaustiva y rigurosa, plantea preguntas un tanto complejas y que se relacionan con uno o más aspectos a la vez. En este sentido, serían más convenientes preguntas más concretas que fueran capaces de responderse mediante “sí” o “no”. Las preguntas, además, pueden estar en un lenguaje similar al del equipo evaluador, pero no para otros actores como el alumnado o el equipo del centro.

Sobre los Ránkings Universitarios no tenemos constancia de las preguntas que se realizan en el cuestionario de prestigio. Sin embargo, dado lo escueto que resulta, pensamos que no se realiza esta diferencia entre lenguaje etic/emic y que directamente se pide el dato que hace referencia al mismo indicador.

Sobre la Evaluación del Mágister (UCM) se pensó que las preguntas del cuestionario dirigido tanto a profesorado como tutorías deberían ser muy precisas y económicas, en el sentido de que se dirigen a un público con una especialización técnica muy alta, así como con relativos pocos recursos adicionales de tiempo como para dedicárselos a una encuesta demasiado farragosa y larga. En este sentido las preguntas del cuestionario se adaptaron a la población la que se dirigían.

6. CONCLUSIONES:

De forma comparativa, mientras que del modelo de la ANECA no se tiene transparencia de la priorización y la selección de las preguntas, estas implican una valoración solo sobre los procesos y algunos productos, están principalmente dirigidas al programa, se aportan estándares relacionados con un gradiente de consecución cualitativa, pero se plantean preguntas demasiado complejas que no se adaptan ni a los públicos ni a las propias personas que realizan la evaluación.

Sobre los Ránkings Universitarios tampoco se tiene constancia sobre la selección de los indicadores, no se plantean tampoco en términos valorativos o interpretativos, algunas preguntas hacen referencia a procesos sociales difícilmente atribuibles del programa, y aunque no se plantean estándares sí se plantean ratios.

Por último, y en relación con los indicadores de evaluación del Mágister en Evaluación (UCM), sobre las preguntas se ha utilizado una fase divergente para su priorización según los criterios establecidos por Cronbach, se plantean preguntas de evaluación que implican una valoración sistémica, se plantean dirigidas únicamente al programa, se utilizan estándares de adecuación, pertinencia, suficiencia y/o calidad, y las preguntas son sencillas.

BIBLIOGRAFÍA:

ALVIRA, F. (1991). Metodología de la evaluación de programas. Madrid, Centro de Investigaciones Sociológicas (CIS).

ANECA (2019). Documento Marco: Evaluación para la renovación de la acreditación de títulos oficiales de Grado y Máster Universitario. Programa ACREDITA. Licencia Creative Commons.

AZIZ, S., MAHMOOD, M., & REHMAN, Z. (2018). "Implementation of CIPP model for quality evaluation at school level: A case study". *Journal of Education and Educational Development*, 5(1), 189-206.

GONZÁLEZ, P. (1986). "Medir en las ciencias sociales". En M. García Ferrando, J. Ibáñez y F. Alvira (coord.), *El análisis de la realidad social*. Madrid, Alianza Universidad Textos, pp.227-286.

LIGERO, J. A. (2017). *Tres métodos de evaluación de programas y servicios: juicios finales sumativos, teoría del cambio y evaluación orientada a los actores implicados*. Madrid, Means Evaluación.

LIGERO, J. A.; PORTA, A.; MUÑOZ, N. y BUSTELO, M. (2019). *Rayuela: un ejercicio de reflexión y comprobación para hacer una evaluación consciente*. Madrid, Máster en Evaluación de Programas y Políticas Públicas (UCM).

MAE-Ministerio de Asuntos Exteriores (2007). *Manual de gestión de evaluaciones de la cooperación española*. Madrid, DGPOLDES.

MAYNE, J. (2011). "Contribution analysis: addressing cause and effect". In Forss, K., Marra, M., & Schwartz, R. (Eds.) *Evaluating the complex: attribution, contribution, and beyond*. Transaction Publishers, 53-96.

MITROFF, I. I., & SILVERS, A. (2010). *Dirty rotten strategies: How we trick ourselves and others into solving the wrong problems precisely*. Stanford, University Press.

NEVO, D. (2006). "Evaluation in education". En SHAW, I. F.; GREENE, J. C. & MARK, M. M. (Eds.), *The Sage Handbook of Evaluation: Policies, Programs and Practices*. London, SAGE, pp.441-460.

RED CIMAS (2015). *Metodologías participativas: sociopraxis para la creatividad social*. Madrid, DEXTRA.

RODRÍGUEZ, R. (2019). *Trabajo Final de Máster*. Madrid, Escuela de Gobierno.

SANCER, H., BATURAY, H. M., & FADDE, P. (2013). "Applying the context, input, process, product evaluation model for evaluation, research, and redesign of an online Master's Program". *The International Review of Research in Open and Distance Learning*, 28(3), 274-290.

SECOLSKY, C., & DENISON, D. B. (EDS.). (2012). *Handbook on measurement, assessment, and evaluation in higher education*. New York, Routledge.

SHAW, I. F.; GREENE, J. C. & MARK, M. M. (2006). *The Sage Handbook of Evaluation: Policies, Programs and Practices*. London, SAGE.

WAGENSBERG, J. (2018). *Sólo se puede tener fe en la duda: Pensamiento concentrado para una realidad dispersa*. Barcelona, Tusquets.

WEISS, C. (1998). *Evaluation*. Upper Saddle River, Prentice Hall.