

La Evaluación Programática (II): elementos claves y posibilidades de aplicación en nuestro sistema evaluativo médico.

Resumen: En este segundo artículo de Doctutor sobre Evaluación Programática (EP), hemos tratado de recopilar los componentes clave de la misma con la idea de sensibilizar y animar a los implicados en la docencia y evaluación médicas sobre cómo estos pueden implementarse en el sistema educativo médico (grado y formación especializada) para mejorarlo de una forma más realista y eficiente. Creemos que en un sistema educativo médico como el español, que se enmarca en un sistema de salud muy competitivo y de muy alta calidad el asumir los principios y las tareas que exige la EP es, no solo factible sino necesario para mantener los estándares del sistema, representando ésta incorporación quizás una oportunidad de mejorarlo que no deberíamos dejar escapar.

Programmatic Evaluation (II): key elements and possibilities of implementation in our medical evaluation system Summary: In this second article of Doctutor on Programmatic Evaluation (PE), we have tried to compile the its key components with the idea of sensitizing and encouraging those involved in medical teaching and evaluation on how these can be implemented into a Medical Education System (degree and specialized training) to improve it in a more realistic and efficient way. We believe that in a medical education system such as the Spanish, which is part of a very competitive and very high quality health system, assuming the principles and tasks required by PE is not only feasible but necessary to maintain the standards of the system, and PE may represent an opportunity to improve it hat we should not miss.



En los lugares en los que la evaluación programática (EP) se ha puesto en marcha en los últimos años (sobre todo en Holanda y en algunos lugares de EEUU, Canadá y Australia) se ha visto que su implementación es costosa y laboriosa, ya que aumenta el tiempo dedicado al feedback, requiere idear mecanismos para apoyar al alumno y guiarlo en el proceso de feedback y aprendizaje

autodirigido, y acordar la toma de decisiones evaluativas entre grupos de expertos para hacer juicios de tipo holístico sobre la capacitación del alumno (1)

Algunas de las descripciones de EP se centran principalmente en el uso del portafolio como medio de captar la evidencia y contribuir así a la toma de decisiones (2,3,4). Sin embargo, se podría aplicar con otros modelos y contextos evaluativos.

Para ver las posibilidades de aplicación en contextos evaluativos particulares, como puede ser el que se utiliza en las escuelas de medicina españolas o el que se establece en la evaluación de nuestros residentes, quizás lo más importante sea el tener en cuenta cuales son los elementos claves de una EP y tratar así de ver en que manera estos elementos pueden ser incorporados a estos contextos evaluativos.

Así, los elementos clave de la EP serían:

1. Crear expectativas claras sobre el aprendizaje que se va a exigir
2. Planificar una selección apropiada de los sistemas evaluativos
3. Centrarse en aquellos alumnos que necesitan atención especial y/o información extra
4. Separar los datos de las decisiones
5. Agregar por atributo (competencia), no por método o tiempo
6. Tomar decisiones conjuntamente, no sobre las evaluaciones de los individuos
7. Promover el intercambio de información y el diálogo en torno a una narrativa y no sobre los resultados numéricos
8. Maximizar las evaluaciones para guiar el aprendizaje.

Mientras que con los primeros tres componentes suelen estar familiarizados todos los implicados en el diseño de cualquier forma de evaluación, los restantes cinco componentes son más específicos de la EP y en ellos precisamente recaen las ventajas de éste tipo de evaluación.

Específicamente, separar los datos de las decisiones ayuda a evitar lo que llamábamos 'fallo al fallar' o equivocarse al calificar (ver el anterior artículo:

<https://www.doctutor.es/2019/02/04/la-evaluacion-programatica-que-es-y-algunas-ventajas-para-la-mejora-del-sistema-evaluativo-e-n-medicina-i/>) (5) los responsables de las evaluaciones individuales solo tienen que dar la información concreta de los resultados de su evaluación al comité para que éste tome las decisiones evaluativas de una forma colectiva, así cualquier decisión difícil recae en un grupo de personas, no en un solo evaluador.

El agregar por atributo y tomar decisiones en base a una información agregada, promueve una compensación más adecuada y reduce el riesgo de una compensación inapropiada, como cuando un alumno o residente con escaso conocimiento es compensado por un buen resultado en unas determinadas habilidades. Esto también puede ayudar a mitigar el "efecto halo", donde el problema es que el evaluador puede juzgar que un estudiante es bueno en un área en base a que ha comprobado que lo es en otra área diferente (6).

El describir la competencia o el desempeño en palabras, en lugar de números, ayuda tanto a los responsables docentes como a los alumnos a dirigir su aprendizaje (7). Reconocemos que un resultado, por ejemplo, de un 5 en una escala evaluativa de 10 lleva mucho menos contenido informativo que una descripción de lo que el estudiante ha hecho bien y de aquello en lo que debe mejorar. Sin embargo, de alguna manera el uso de números implica una mayor objetividad, algo que, sin embargo, no está respaldado por la evidencia. El uso de números también hace que todo sea mucho más fácil simplemente hay que sumarlos: el total de la ECOE, más

los exámenes de preguntas de elección múltiple determina el nivel que tiene el alumno, lo que aumenta el riesgo de una compensación inapropiada (8). Se ha demostrado que el uso de una narrativa, sin la necesidad de un portafolio, para describir áreas que un alumno pueda necesitar trabajar más, tiene un impacto positivo sobre la capacidad del sistema para detectar y actuar sobre los alumnos que muestran malos comportamientos profesionales (9).

Problemas relacionados con la implementación

Muchos modelos requieren el uso de un portafolio (2,3,4) y cambios a gran escala en los métodos de evaluación. Cualquier sistema de evaluación mejora si éste tiene definidos sus objetivos, si existe compromiso del personal con suficiente experiencia en la evaluación y si hay un control centralizado de la misma (10). Los siete elementos claves antes descritos se pueden incluso resumir en los siguientes componentes fundamentales para la aplicación de un EP:

1. Blueprinting 2. Recolección de datos 3. Toma de decisiones 4. Apoyo del personal 5. Evaluación Blueprinting

En nuestro país son muy escasas las instituciones educativas que planifican la evaluación de las diferentes competencias y materias mediante "blueprinting". El mismo clarifica el propósito, o propósitos, de toda evaluación. En general, hay tres propósitos principales: 1) guiar el aprendizaje; 2) informar sobre la mejora de la calidad de desarrollo curricular; y/o 3) tomar decisiones importantes sobre los alumnos. El primer propósito se centra en el aprendizaje, que suele ser donde los estudiantes tienen dificultades. El segundo propósito se centra en cómo podría cambiarse un plan de estudios para abordar las dificultades que tienen muchos estudiantes, ¿De qué manera un plan de estudios impulsa comportamientos de aprendizaje? El tercer propósito incluye certificar si alguien está preparado para pasar a la siguiente etapa formativa (curso, año de residencia, titulación, etc). A veces estos propósitos evaluativos generales se superponen, pero en otras veces el tratar de conseguir uno puede afectar negativamente a otro. Por ejemplo, Si la evaluación se diseña para guiar el aprendizaje, a los estudiantes se les recomendará que reconozcan abiertamente sus debilidades, pero si la evaluación es para certificar la competencia, lo normal es que los alumnos oculten sus fallos. Otro ejemplo podría ser cuando el plan de estudios y sus objetivos de evaluación tienen como meta promover la colaboración, y sin embargo, la evaluación también se utiliza para selección, aquí los aspectos de colaboración y los de competencia entran en conflicto. La clave en estos casos es ser explícito sobre los propósitos con los que se plantea la evaluación y sobre cualquier compromiso que haya que asumir. Lo que implica tener claro las competencias (atributos) en las que los alumnos deben ser evaluados. Tales atributos o competencias podría incluir, por ejemplo, habilidades de trabajo en equipo, apuntalar conocimientos, habilidades de procedimiento o comunicación interpersonal. Una vez hecho esto, se debe elaborar un plan de evaluación, donde hay que ser claro sobre cómo cada herramienta evaluativa contribuye a cada atributo. Esto se suele documentar como una cuadrícula, donde las filas son los atributos y las columnas son las herramientas de evaluación. Las celdas de la cuadrícula indican qué herramientas están diseñadas para evaluar qué atributos. Tradicionalmente, las decisiones se toman mirando los resultados de cada episodio evaluativo o de un período de tiempo corto de evaluaciones, como dentro de un único rotatorio clínico (columnas), mientras que en la EP, éstas se tomarían de los resultados para cada atributo (filas).

Recopilación de datos

Para decidir si un estudiante o residente ha alcanzado un nivel de rendimiento satisfactorio en cada atributo, se debe tener una colección de evidencias. Dicha evidencia comprende el resultados de cada episodio de evaluación. En EP, las decisiones sobre datos se separan de la recopilación de los datos. Esto significa que los datos que se recopilan necesitan enmarcarse de forma que expliquen dónde está un alumno en su aprendizaje. En otras palabras, idealmente deben documentar su fortalezas y debilidades, en lugar de una decisión. Los datos también deben recopilarse de manera que sea fácil clasificarlos y sintetizarlos, para que puedan presentarse de manera significativa a la hora de la toma de decisiones. También deberían ser presentados de nuevo al alumno de manera que sea fácil para ellos decidir hacia dónde enfocar su aprendizaje. Esta es un área donde puede haber muchos mecanismos innovadores para recolectar tales datos. La toma de decisiones debería ser idealmente por atributo en lugar de por el método de evaluación; por ejemplo, saber si alcanzó el alumno un nivel satisfactorio en comunicación, en lugar de si ha superado o no un Mini-CEX. Las implicaciones de esto para la recopilación de datos son que los registros de rendimiento deberán coincidir con el atributo en cuestión (p. ej. comunicación), no solo con las herramientas utilizadas para evaluar ese rendimiento (por ejemplo, Mini-CEX, entrevistas a PS, etc). La tecnología de la información puede ser muy útil aquí.

Toma de Decisiones

Hay dos grandes tipos de decisiones a tomar: decisiones de menor trascendencia que guían el aprendizaje y aquellas en las que el aspirante se juega el progresar a un nivel superior o la certificación. El proceso de recogida de la evidencia es similar en ambos casos, pero la fuerza y el peso de la evidencia debe ser mayor en las segundas.

Una vez que los datos se han recopilado para que se puedan presentar de manera útil, deben de ser sintetizados y analizados. La información relacionada con cada atributo se debe considerar colectivamente para tomar una decisión sobre ese atributo en cada alumno. El proceso se repite después para el siguiente atributo antes de ofrecer una decisión final en cada alumno. Esto representa un trabajo importante, pero para muchos estudiantes la decisión lo es aún más ya que aunque muchos superarán claramente todos los requisitos (es decir, pasarán o aprobarán) mientras que otros pueden claramente quedarse cortos (es decir, suspenderán y no pasarán). Para estos alumnos, la presentación y decisión resultante de la información será seguramente rápida, sin embargo, siempre existirá un grupo de alumnos en los que las decisiones serán más difíciles, particularmente donde hay incertidumbre sobre si existe evidencia suficiente para determinar la capacidad de un alumno en un atributo particular, donde esta evidencia es contradictoria, o donde un alumno se muestra claramente más fuerte en un atributo pero más flojo en otro. Aquí es donde se necesitará el juicio de los expertos. Como en muchas otras áreas del ámbito de la toma de decisiones, tales decisiones es mejor que la tomen un grupo de expertos (11). Como tal, para decisiones importantes, la mayoría de los sistemas de EP utilizan los llamados "Comités de Progresión". Las decisiones secundarias o de "menor riesgo" suelen ser tomadas por un supervisor, un mentor y/o, por supuesto, el alumno.

Las formas de mejorar la capacidad para defender la solidez de las decisiones incluyen la capacitación, el uso de información de tipo narrativa y procesos explícitos de apelación. Este proceso será asistido al tener un marco de referencia común para todas las evaluaciones (por ejemplo, por dominio competencial o atributo) pero también debe tener en cuenta factores que impactan en la toma de decisiones grupales, como el poder de cada uno de los miembros del grupo, la existencia o no de veto (o pérdida de veto) y, por lo tanto, hasta qué punto las opiniones individuales pueden tener un peso inapropiado. Todos estos procesos tienen similitudes con la investigación cualitativa y con la toma de decisiones clínicas. La credibilidad y la fiabilidad en la investigación cualitativa se puede lograr mediante procesos de triangulación, compromisos mantenidos, verificación de miembros, seguimiento con auditorías,(12). Se pueden usar procesos similares en las decisiones de evaluación normal (12). Asimismo, los profesionales de la salud están acostumbrados a tomar decisiones en base a una información dispar y múltiple (p. ej., evaluar diferentes pruebas cardiacas, como: ruidos, presión venosa presencia o no de edema, para determinar la existencia o ausencia de insuficiencia cardíaca). Se han hecho otras comparaciones entre este tipo de toma de decisiones en evaluación y la que se hace en la clínica (13).

Apoyo del personal

El incorporar la EP, supondrá para muchos docentes y personal académico, un cambio de paradigma (14). Esto inevitablemente significa que muchos de ellos necesitarán formación, orientación y/o capacitación. Parte de esta formación incluirá garantizar que haya una buena documentación en torno a la justificación del programa de evaluación de cada institución y de cómo encajan sus diversos componentes (11,15). Personal y alumnos deberían de estar particularmente interesados e implicados en todo el proceso de toma de decisiones: ¿Qué evidencia queremos? ¿Cómo la sintetizamos? ¿Cómo garantizamos la credibilidad y la fiabilidad? En nuestro medio esto supondrá pedirle al personal que cambie lo que hasta ahora vienen haciendo, además de plantear problemas de coste y rentabilidad (1). Si bien, en este sentido, se puede requerir un mayor esfuerzo en algunas áreas, seguramente habrá que redistribuir y economizar recursos en otras. Tal redistribución podría orientarse a conseguir un mayor uso de datos de observaciones recopiladas rutinariamente (como son las observaciones hechas en el lugar de trabajo) y un menor uso de exámenes caros (como por ejemplo las ECOE) (1).

Evaluación

Cualquier sistema de evaluación necesita procesos integrados de mejora de la calidad, es decir precisa a su vez ser evaluado. También se debe de investigar más al respecto. Mientras se empieza a tener alguna evidencia sobre el impacto (16) y efectividad de

la EP (17), se precisará conocer más, sobre la efectividad de los métodos de recolección de información y decisiones a tomar, y también sobre como la EP influye en los comportamientos de los alumnos y docentes.

Conclusiones

La EP representa un cambio de paradigma, especialmente en un contexto educativo (no solo evaluativo) como el nuestro que precisa cambios profundos y rigor de cara a los propios alumnos, a los profesionales pero también para rendir cuentas socialmente en una materia sensible. El asumir sus principios y métodos implicaría una reformulación de los enfoques educativos que en nuestro caso es muy necesaria y que debe llevar a un cambio de mentalidad en la docencia y evaluación. Su asunción por lo tanto supone un auténtico desafío tanto para el sistema como para sus protagonistas, docentes y alumnos, donde todos nos jugamos mucho. Sin embargo, existe el riesgo de que sea visto tanto como capaz de resolver todos los problemas, como lo contrario, que no resolverá ninguno, y que incluso causará más. Es probable que la realidad esté en el punto medio. Además, como con cualquier cambio de paradigma, están los puristas que sienten que solo se puede hacer de una manera, contrarrestada por otros que ven la EP más como una forma de trabajar mediante la cual hay margen para la flexibilidad en su desarrollo, implementación, y evolución.

En este segundo artículo de Doctutor sobre EP, hemos tratado de recopilar los componentes clave de la misma con la idea de sensibilizar y animar a los implicados en la docencia y evaluación médicas sobre cómo estos pueden relacionarse con los objetivos de aprendizaje de una forma más realista y eficiente. Creemos que en un sistema educativo médico como el español, que se enmarca en un sistema de salud muy competitivo y de muy alta calidad el asumir los principios y las tareas que exige la EP es, no solo factible sino necesario para mantener los estándares del sistema.

Referencias

1. van der Vleuten CPM, Heeneman S. On the issue of costs in programmatic assessment. *Perspect Med Educ*. 2016;5(5):303?307.
2. Dannefer EF, Henson LC. The portfolio approach to competency-based assessment at the Cleveland Clinic Lerner College of Medicine. *Acad Med*. 2007;82(5):493?502.
3. Fishleder AJ, Henson LC, Hull AL. Cleveland Clinic Lerner College of Medicine: an innovative approach to medical education and the training of physician investigators. *Acad Med*. 2007;82(4):390?396.
4. Roberts C, Shadbolt N, Clark T, Simpson P. The reliability and validity of a portfolio designed as a programmatic assessment of performance in an integrated clinical placement. *BMC Med Educ*. 2014; 14(1):197.
5. Wilkinson TJ, Wade WB. Problems with using a supervisor's report as a form of summative assessment. *Postgrad Med J*. 2007;83(981):504?506.
6. Nisbett RE, Wilson TD. The halo effect: evidence for unconscious alteration of judgments. *J Pers Soc Psychol*. 1977;35(4):250?256.
7. Weller JM, Misur M, Nicolson S, et al. Can I leave the theatre? A key to more reliable workplace-based assessment. *Surv Anesthesiol*. 2015; 59(4):169.
8. Tweed M. Passing assessment should not just be jumping hurdles. *Focus on Health Professional Education: A Multi-disciplinary Journal*. 2010;11(3):85?89.
9. Wilkinson TJ, Tweed MJ, Egan TG, et al. Joining the dots: conditional pass and programmatic assessment enhances recognition of problems with professionalism and factors hampering student progress. *BMC Med Educ*. 2011;11(1):29.
10. Timmerman AA, Dijkstra J. A practical approach to programmatic assessment design. *Adv Health Sci Educ Theory Pract*.

2017;22(5):1169?1182.

11. Oudkerk Pool A, Govaerts MJB, Jaarsma DADC, Driessen EW. From aggregation to interpretation: how assessors judge complex data in a competency-based portfolio. *Adv Health Sci Educ Theory Pract*. Epub 2017 Oct 14.
12. Driessen E, van Der Vleuten C, Schuwirth L, van Tartwijk J, Vermunt J. The use of qualitative research criteria for portfolio assessment as an alternative to reliability evaluation: a case study. *Med Educ*. 2005;39(2):214?220.
13. Schuwirth L, van der Vleuten C, Durning SJ. What programmatic assessment in medical education can learn from healthcare. *Perspect Med Educ*. 2017;6(4):211?215.
14. Ellis R, Hogard E. Programmatic assessment: a paradigm shift in medical education. *AISHE-J: The All Ireland Journal of Teaching and Learning in Higher Education*. 2016;8(3):29501?29515.
15. Dijkstra J, Van der Vleuten CPM, Schuwirth LWT. A new framework for designing programmes of assessment. *Adv Health Sci Educ Theory Pract*. 2010;15(3):379?393.
16. Heeneman S, Oudkerk Pool A, Schuwirth LWT, van der Vleuten CPM, Driessen EW. The impact of programmatic assessment on student learning: theory versus practice. *Med Educ*. 2015;49(5):487?498.
17. Schuwirth LWT, van der Vleuten CPM. Programmatic assessment and Kane's validity perspective. *Med Educ*. 2012;46(1):38?48.