

# LA EVALUACION DE LAS COMPETENCIAS

*Valea Pérez, Angel; González Arce, María Luz. Universidad del País Vasco*

## **Resumen**

Recientemente hemos vuelto a reflexionar sobre el amplio y confuso concepto de “competencia” y sobre la estrategia y la táctica de su evaluación, ya que, si en lo sucesivo hay que seguir realizando este tipo de acciones (homologación de un título, incluso con las evaluaciones que se consideren oportunas) será difícil hacerlo unos si no se ha planificado de forma uniforme por competencias en el país de origen e incluso, aunque se haya hecho, ¿cómo se puede garantizar la validez externa y universal de esas competencias?

## **Introducción**

En algunos trabajos anteriores ya hemos tenido ocasión de tratar aspectos relacionados con la evaluación, donde se evidenciaba una variada problemática en asunto tan sensible como es valorar la formación de personas. Entre un mar de opiniones y experiencias, descubrimos algunos autores que dicen que *“la evaluación es quizás el más importante de los procesos involucrados en la formación técnica y profesional (Linn y Gronlund; Popham; Scallon; Gronlund y Cameron; Shay; Tardif; Gullickson)”* con los que, modestamente, no podemos coincidir ya que un medio de comprobación no parece que debería ser el proceso más importante, aunque ciertamente parece necesario considerar que deberá existir coherencia entre el objeto a evaluar y el procedimiento seleccionado para ello. Por tanto, si no se tiene claro aún el concepto de competencia (dentro del ámbito universitario) difícilmente se puede adecuar el procedimiento de evaluación.

Ciñéndonos a la universidad española (y por lo que conocemos, a otras muchas universidades extranjeras) si la mayoría de los profesores han dedicado toda su vida a la llamada “carrera docente” sin haber siquiera “visitado” una actividad profesional o una empresa privada; es dudoso que pueda evaluar competencias profesionales, *“...destrezas conocimientos y actitudes que se aplican en situaciones reales....”*

## **Marco teórico**

Decía, el filósofo español Antonio Machado que *“el camino se hace al andar”*, y mucho antes lo había dicho uno de los padres de la filosofía aristotélica *“... Nos volvemos justos realizando actos de justicia...”*. Si hacemos historia, algunos autores (Madaus y O’Dwyer) remontan las pruebas de desempeño nada menos que a la dinastía Han en China. En la Edad Media se utilizaron evaluaciones similares de competencias en las organizaciones gremiales. En las primeras universidades europeas se utilizaban discusiones orales, de contenido filosófico o teológico, en latín, para la valoración de las competencias de los estudiantes, entre las que incluían aspectos del desempeño (p.ej. ... *the ability to present it in eloquent form, and a tacit conformity to orthodoxy...*)

tan habituales en el ámbito anglosajón . Mucho más cercano en el tiempo se sitúan los Assessment Center (Centros de Evaluación) que utilizan diferentes herramientas y simulaciones para evaluar competencias complejas. En el mismo ámbito se encuentran las Oficinas de Selección de Personal, el examen para conducir vehículos que hace la Dirección General de Tráfico o los exámenes musicales en los Conservatorios de Música, Ballet Nacional, etc. En todos ellos se le pide al estudiante que demuestre el nivel de competencia en cada una de las especialidades de desempeño en situaciones casi reales.

En contextos profesionales, cuando es necesario demostrar la competencia profesional, se desarrollan procedimientos para evaluarlos. Así p.ej. para inspectores de Ensayos No Destructivos, es necesario demostrar periódicamente la competencia en cada una de las técnicas para poder seguir ejerciendo como tal; del mismo modo en algunos estados de los EE.UU se pide una certificación periódica a médicos (United States Medical Licensure Examination) y a abogados (American Bar Association) por la cual acreditan sus niveles de competencia para poder seguir ejerciendo en situaciones casi reales. Algunas grandes compañías multinacionales también tienen sus evaluaciones de desempeño para poder avanzar en su carrera profesional y, en cierta medida, la Universidad se alinea, aunque con matices, en esta dirección. Actualmente, ante las elecciones en España se vuelven a oír tambores acerca de evaluaciones periódicas a los profesores de todos los niveles (probablemente como argumento político).

### ***Metodología***

Parece evidente que la medida de una competencia pasa por una prueba que conduzca a la valoración del dominio de dicha competencia en situaciones reales. Establecida la competencia y admitido que la prueba es capaz de evaluarla, llegaríamos a la conclusión de que ésta define un determinado grado de “saber hacer” que es preciso definir.

#### 1. Medida de las competencias

Desde mediados de los ochenta, las herramientas de evaluación han ido cambiando progresivamente. Estos cambios, se supone, que vienen motivados por la creencia de que las pruebas convencionales (papel y lápiz) o las de respuesta múltiple no son adecuadas para medir habilidades complejas, ni para medir las destrezas para el aprendizaje continuo. Pero definir una realidad tan compleja es confuso y más aún si, como se preconiza, deben medirse en situaciones “casi reales”. En el ámbito educativo la evaluación de las competencias supondría una aproximación al futuro desempeño que deberían ser capaces de realizar los futuros profesionales y actuales alumnos. La demostración del conocimiento aplicado a posibles situaciones futuras de la vida real solo puede realizarse a través de simulaciones de la vida real, tal como destaca en la definición del National Council on Measurement in Education (NCME). En la misma línea, autores tales como Stiggins; Ruiz-Primo y Shavelson, definen la evaluación de competencias como un sistema de evaluación que incluye: a) los desempeños que quieren evaluarse b) un conjunto de tareas que simulan el desempeño a evaluar c) unas respuestas a esas tareas que simulan el desempeño. Sin embargo, en esta

evaluación de competencias basadas en una simulación de tareas existen dos defectos subyacentes de modelo. El primero que la evaluación de la competencia se realiza en un ámbito exclusivamente académico y no real y la segunda es que no se trataría de demostrar las competencias sino de contestar una simulación.

### **Resultados**

#### 1. Experiencias psicométricas de las pruebas de competencias en educación

Para que las informaciones obtenidas en el sistema de evaluación puedan transformarse en evidencias, necesitamos: a) Un modelo de representación del conocimiento que suponen las competencias a evaluar. b) Un sistema de observación del desempeño que sea coherente con el modelo de representación del conocimiento. c) Método de interpretación de las respuestas dadas al sistema de observación. Estos tres elementos constituirían lo que, en este caso, se identificaría con el “triángulo de la evaluación” (National Research Council, 2001). Las pruebas de evaluación de competencias, debe cumplir con los criterios psicométricos (Kane). Algunas de las características psicométricas que deben estudiarse en la evaluación de competencias son: A) modelo de medida B) Fiabilidad de las mediciones C) Capacidad de discriminar entre evaluados y sistematización de la medida D) Validez de las medidas.

A) Modelos de medida: Como ya se ha dejado escrito, el proceso de evaluación consiste en establecer una relación, cuantitativa, entre las respuestas dadas por los alumnos al sistema de evaluación y la propia competencia a valorar. Esta relación se establece a través de un modelo estadístico que se llama modelo de medida. En el caso de querer evaluar competencias, las informaciones no pueden limitarse a una respuesta correcta/ incorrecta sino que dan respuestas polinómicas, lo que hace que “a priori” no sean adecuados los modelos basados en la Teoría de Respuesta al Item (TRI). Aunque es cierto que se han desarrollado modelos TRI politómicos (Wu, Adams, Wilson y otros) aún no se ha estudiado suficientemente la adecuación de estos modelos a los sistemas de evaluación de competencias (Reckase, Gibbons, Bock, Hedeker, Weiss, Kupfer, Grochocinsky, Stover, etc.).

B) Fiabilidad: La medida de la fiabilidad en la evaluación de competencias necesita un enfoque más amplio que el dado por la Teoría Clásica de los Test (TCT). Una extensión de la Teoría Clásica de los Test es la Teoría de la Generalizabilidad (TG) de Gronbach, Gleser, Nanda y Rajaratnam, que utiliza el Análisis de los Componentes de la Varianza para estimar simultáneamente los efectos de las diferentes fuentes de error sobre las puntuaciones (Martinez Arias, 2010). Para tratar de resolver este problema Miller y Linn (2000) sugieren dos soluciones: a) incrementar el número de tareas de la prueba b) definir de una forma más precisa la competencia a medir y la propia tarea de evaluación. La segunda cuestión hace referencia a la afectación del factor humano en la evaluación de las competencias. El nivel de correlación entre correctores diferentes, en un estudio de Lane y Stone (2006) ha demostrado ser inadmisiblemente pequeño y con desviaciones altas (entre 0.33 y 0.91) para una evaluación en escritura. La fiabilidad entre correctores depende también del carácter cuantitativo de la prueba, así se encuentra mayor consistencia en pruebas de matemáticas que en una de escritura, por ejemplo (Shavelson, Baxter y Gao, 1993).

D) La validez de las medidas: En los Standards for Educational and Psychological Test (AERA et al, 1999) se establece la validez como “ *el grado en que la evidencia y la teoría apoyan las interpretaciones que se van a hacer de las puntuaciones obtenidas en las pruebas de medida...*”. Al tratar la validez deberíamos separar al menos tres tipos de validez: de contenido, sustantiva y externa. *La validez de contenido* se refiere a la representatividad del contenido de la evaluación. Aquí hay dos aspectos relevantes anotados por Messick (1996); por un lado la posible “infra-representatividad” debido a que las pruebas de competencias suelen incluir menor cantidad de estímulos que las pruebas tradicionales tipo papel y lápiz, y en segundo lugar lo que llama el “riesgo de varianza irrelevante” debido a fuentes de error asociadas con la tendencia de los correctores humanos a fijarse en aspectos poco relevantes de las respuestas. En cuanto a la *validez sustantiva*, como capacidad para medir habilidades complejas y de alto nivel cognitivo, digamos que son pocas las investigaciones que avalen esta validez sustantiva y Martínez Arias (2010) muestra que los resultados obtenidos son poco consistentes. La *validez externa* se refiere a las relaciones de las puntuaciones de un test con variables externas. Miller y Linn, 2000, dicen que las pruebas de competencias no suelen estar basadas en teorías bien definidas. El procedimiento para obtener evidencias de validez externa es el estudio de correlaciones según las condiciones de contorno impuestas al modelo. Martínez Arias (2010) dice que hay pocos trabajos sobre estas evidencias en educación.

### **Conclusiones**

Bajo el punto de vista de la propia evaluación hay serias dudas sobre los imprescindibles requisitos psicométricos y, en cualquier caso, parece haber coincidencia en que si se van a seguir aplicando, urge una mayor investigación para aportar solidez técnica a los sistemas de medida de las competencias.

La inconsistencia de valoraciones entre diferentes examinadores y la del mismo examinador en diferentes momentos de tiempo, son fuentes de error incontrolable e inadmisibles.

### **Referencias**

- American Educational Research Association, American Psychological Association & National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing*. Washington, DC: AERA.
- Brennan, R.; (2000) Performance Assessment from the perspective of the Generalizability theory. *Applied Psychological Measurement*, 24, 339-353
- Gibbons, R.D.; Bock, R.D.; Hedeker, D.; Weiss, D.J.; Segawa, E.; Bhaumik, D.; Kupfer, D.J.; Frank, ED.; Grochocinsky, V.J.; Stover, A.; (2007); Full information item bifactor analysis of graded response data. *Applied Psychological Measurement*, 31, 4-19
- Kane, M.T.; (2004); Certification Testing as an illustration on argument-based validation. *Measurement: Interdisciplinary Research and Perspectives*, 2, 13-170

- Lane, S.; Parke, C.S.; Stone, C.A.; (2002). The impact of a state performance-based assessment and accountability program on mathematics instruction and student learning: Evidence from survey data and school performance. *Educational Assessment*, 8, 279-315
- Martinez Arias, R.; (2010); La evaluación del desempeño. *Papeles del Psicólogo*, 31(1), 85-96
- Messick, S.; (1996); Validity of performance assessments. In *Technical issues in large-scale performance assessment* (Ed. G. Phillips). Wash. National Center for Education Statistics, 1-18
- Miller, D.M.; Linn, R.L.; (2000) Validation of performance assessment. *Applied Psychological Measurement*, 24, 367-378
- Palm, T.; (2008) Performance Assessment and authentic assessment: a conceptual analysis of the literature. *Practical Assessment, Research and Evaluation*, 13 (4)
- Shavelson, R.J.; Baxter, G.P.; Gao, X.; (1993); Sampling variability of performance assessments. *Journal of Educational Measurement*, 30, 215-232.
- Van der Vleuten, C.; Swanson, D.; (1990); Assessment of clinical skill with standardized patients: State of the art. *Teaching and Learning in Medicine*, 2, 58-76