

EVALUACIÓN DE COMPETENCIAS PARA LA CERTIFICACIÓN DE PROFESIONALES DEL SISTEMA DE JUSTICIA PENAL EN MÉXICO

ASSESSMENT IN PROFESSIONAL SKILLS CERTIFICATION PROGRAM TO CRIMINAL JUSTICE SYSTEM IN MEXICO

Leyva-Barajas, Yolanda E.

Fecha de recepción y de aceptación: 3 de noviembre de 2022, 19 de diciembre de 2022

Resumen: En este artículo se presenta la experiencia del diseño, desarrollo e implementación de un modelo de evaluación de competencias profesionales para un programa de certificación para la justicia en México (CEJUME). Se describen las estrategias implementadas acordes con las tendencias actuales a nivel teórico y metodológico para enfrentar los principales desafíos de una evaluación de desempeño basada en casos reales. Consideramos que este artículo aporta información valiosa y útil para el desarrollo de programas de evaluación de desempeño profesional en el contexto real en el cual se produce. Si bien fue diseñado para la evaluación de profesionales del Derecho que operan el sistema de justicia penal en México (SJP), los desafíos son similares a los de otros campos profesionales. Se enfatiza la importancia de llevar a cabo estudios de investigación a partir de los resultados de las primeras evaluaciones para validar los argumentos de que este modelo de evaluación de competencias, promueve la reflexión sobre la práctica profesional favoreciendo su carácter formativo. Los hallazgos aquí reportados nos han permitido recuperar evidencia para sustentar decisiones de mejora del modelo y una mayor confianza en los resultados de la evaluación para los usos previstos a nivel de las personas evaluadas y de las instituciones de justicia en las que trabajan.

Palabras clave: Evaluación auténtica, competencias profesionales, portafolios de evaluación, evaluación de desempeño, modelos cognitivos, validez

Abstract: This paper presents the experience of designing, developing and implementing a professional competency assessment model for a certification program for justice in Mexico (CEJUME). The evaluation model includes current trends both theoretical and methodological to face the main challenges that represent a fair evaluation based in real cases. While it was designed for the evaluation of legal professionals operating the criminal justice system in Mexico (SJP), the challenges are similar to those in other professional fields. The importance of carrying out research studies based on the results of the first evaluations is emphasized in order to validate the main arguments. Some of this arguments are that the portfolios evaluation promotes reflection on professional practice favoring its formative character. The findings reported here have allowed us to retrieve evidence to improve the model and greater confidence in the results of the evaluation for the intended uses at the level of the people evaluated and the justice institutions in which they work.

Keywords: Authentic evaluation, professional skills, evaluation portfolios, performance evaluation, cognitive models, validity

1. INTRODUCCIÓN

La evaluación de competencias en cualquier campo profesional, requiere de la medición de constructos complejos, retomando para ello una base conceptual que permita la creación de nuevas medidas de progreso y ejecución. Los desafíos que se enfrentan actualmente en el ámbito de la evaluación de competencias profesionales demandan una comprensión profunda de los conceptos clave de diversos campos disciplinares, que se traduzcan en habilidades para aplicar este conocimiento en la explicación de fenómenos y la solución de problemas de la vida real (Erickson & Oliveri, 2016; Pellegrino & Hilton, 2012; NRC, 2013).

Si bien desde la década pasada se han reportado importantes resultados en esta línea de trabajo en diversos reportes del National Research Council (NRC) y se ha presentado una nueva generación de estándares en aprendizaje de las ciencias que reflejan esta nueva visión (NRC, 2012; NRC, 2013), son muy pocos y aislados los esfuerzos a nivel de la evaluación de competencias profesionales en el campo laboral, principalmente en países iberoamericanos.

Si realmente queremos mejorar la práctica de profesionales en servicio, tenemos que desarrollar programas de evaluación con carácter formativo que recuperen evidencia directa del desempeño en una gran variedad de contextos y circunstancias. Esto implica pasar de la medición de constructos simples que reflejan la memorización de hechos, procedimientos o principios, a la medición de constructos complejos que se componen de recursos cognitivos, interpersonales e intrapersonales centrados en la aplicación y transferencia del conocimiento en una gran diversidad de contextos laborales y bajo diferentes condiciones institucionales (Erickson & Oliveri, 2016; NRC, 2013a; Pellegrino & Hilton, 2012).

Los desafíos asociados con este cambio en la evaluación son múltiples, requieren en primera instancia, entender que no es posible evaluar la aplicación del conocimiento mediante constructos simples; que resulta indispensable introducir situaciones basadas en hechos reales que reflejen variantes que provienen de múltiples dimensiones del contexto en el que suceden los fenómenos que intentamos explicar y los problemas que necesitamos resolver (Pellegrino, Wilson, Koenig, & Beatty, 2014).

Desafíos y tendencias para la evaluación de competencias

La evaluación de constructos complejos requiere el desarrollo de un marco teórico basado en modelos cognitivos que sirva de fundamento para la definición del dominio a evaluar (Bussemeyer y Diederich, 2010). El primer paso consiste en la definición del dominio a evaluar y constituye el primer desafío considerando que es la base en la que se sustenta la validez de todo modelo de evaluación desarrollado.

El siguiente desafío se concentra en la elaboración de tareas evaluativas para la integración de un Portafolios de evaluación mediante el cual se pueda documentar la práctica con artefactos que constituyan evidencia directa de desempeño y que favorezca la reflexión en torno a las decisiones y diversas acciones desarrolladas en el cumplimiento de las funciones profesionales. El portafolios de evaluación ha resultado ser una estrategia útil para hacer evaluación auténtica basada en la selección de casos reales atendidos por las personas evaluadas (Barbera, 2005; Darling-Hammond, Falk and Ancess, 2017).

Otro aspecto importante en una evaluación de este tipo es la representatividad del contexto en que se produce la práctica profesional. Las pruebas basadas en casos permiten evaluar el desempeño profesional en una gran diversidad de contextos, principalmente si son las propias personas evaluadas quienes seleccionan los casos que usarán para demostrar su capacidad en las distintas etapas del proceso de solución de problemas. Este tipo de pruebas ha dado buenos resultados en la evaluación de competencias médicas. También en la evaluación de docentes y se les denomina “pruebas de desem-

peño óptimo” ya que los propios docentes deciden el tema, asignatura, grado, etc., sobre el que quieren mostrar sus habilidades de planeación, intervención y evaluación educativa en el contexto de la integración de sus portafolios de evaluación. El modelo es muy conveniente también en el ámbito jurídico porque las y los abogados resuelven casos específicos en diversos campos de su ejercicio profesional.

Cuando se decide hacer evaluación auténtica mediante la integración de portafolios se presentan nuevos desafíos a nivel del desarrollo de tareas evaluativas y de la construcción de rúbricas de calificación para el análisis e interpretación de resultados, con el propósito de valorar las cualidades que deben mostrar las respuestas y artefactos seleccionados en la integración del portafolios. La elaboración de los descriptores de desempeño de las rúbricas debe sustentarse en el modelo cognitivo desarrollado en el marco teórico para producir interpretaciones válidas del desempeño observado y favorecer procesos de retroalimentación útiles y significativos para que la evaluación cumpla su carácter formativo.

De acuerdo con los estándares de evaluación de la APA, AERA, NCME (2014), un aspecto clave en el trabajo de diseño y desarrollo de una evaluación, es la participación de las diversas audiencias implicadas desde la planeación del modelo para definir los propósitos, delimitar los usos y acordar los mecanismos y estrategias que aseguren una evaluación equitativa y justa para todas las personas evaluadas. Este requerimiento implica un desafío de tipo político porque requiere el despliegue de una serie de mecanismos y estrategias de negociación con toda las partes implicadas y de comunicación y difusión de los principios básicos del modelo que se construya, orientados a garantizar la equidad, el reconocimiento de la diversidad de las personas evaluadas y de los contextos en los que trabajan y la participación de la población objetivo en todas las fases del proceso evaluativo, desde la construcción misma del modelo, hasta los procesos de interpretación y uso de los resultados. Las negociaciones son fundamentales para dar credibilidad y viabilidad a la evaluación.

La evaluación con fines formativos y de certificación requiere la elección de un modelo de calificación y un método para el establecimiento de puntos de corte para la emisión de un dictamen en términos del desempeño observado. La emisión de un dictamen basado en los resultados de pruebas que evalúan constructos complejos constituye otro desafío debido al reconocimiento de la diversidad y riqueza de las observaciones obtenidas a partir de la integración de respuestas y artefactos de la práctica en una gran diversidad de contextos y condiciones.

El desafío obliga a definir un modelo de calificación que permita discriminar de manera adecuada quiénes cumplen con los estándares de competencia establecidos y quiénes requieren apoyo para lograrlos. El concepto de estándares tiene matices relacionados con el tipo de instrumento al que se refieran. De acuerdo con Jornet, González-Such, Suárez y Perales (2011) en aproximaciones de tipo cualitativo, actúan como un argumento de referencia para poder explicitar los criterios de evaluación que faciliten la comparación de la ejecución que muestra una persona con un referente que facilite la evaluación de la calidad del aprendizaje observado. La definición de estos estándares de desempeño se construyen las rúbricas cuidando que los descriptores de desempeño no produzcan sesgos y se orienten a valorar la pertinencia de las acciones y decisiones tomadas, al contexto y a las características de los intervinientes.

Finalmente tenemos el desafío de establecer procesos de validación que se centren en las demandas, los supuestos clave y las inferencias que están ligadas a las puntuaciones de una evaluación con sus interpretaciones y usos previstos. De acuerdo con Cook, Brydges, Ginsburg y Hatala (2015), un proceso serio y sistemático de validación implica articular las demandas y supuestos asociados, con las decisiones propuestas como la interpretación y el uso de argumentos, probando de manera empírica estos supuestos e inferencias y organizando la evidencia en argumentos de validez coherentes.

En otras palabras, los procesos de validación implican el desarrollo de líneas de investigación que permitan recabar evidencia que apoye cada una de las inferencias, enfocándose en los supuestos más cuestionables de la cadena de inferencias (Kane, 2006). Las conjeturas o supuestos clave, así como la evidencia que se requiere recabar, dependen en gran medida del uso intencionado o asociado a las decisiones que se van a tomar.

La importancia de llevar a cabo este tipo de estudios de investigación, además de proporcionar evidencia de validez, es que contribuyen a la comprensión del desempeño en la solución de casos y problemas propios de determinados campos profesionales y permiten indagar sobre las consecuencias no intencionadas de las evaluaciones para evitar cualquier afectación no prevista a la población objetivo.

Programa de Certificación para la justicia en México

En este espacio se describen las estrategias implementadas para enfrentar los desafíos presentados en cada fase del proceso evaluativo. Es importante enfatizar que las estrategias aquí presentadas se incorporaron como elementos del modelo de evaluación de competencias desarrollado en el marco del proyecto de certificación para la justicia en México (CEJUME), construido en el CEEAD con el subsidio otorgado por el Departamento de Estado de los Estados Unidos de América, por conducto de la Oficina Internacional de Asuntos Antinarcóticos y Procuración de Justicia (INL).

Todo el trabajo se llevó a cabo bajo un marco de criterios y directrices para construir una evaluación justa, válida y útil (CEEAD, 2019) y en estándares técnicos internacionales en materia de evaluación educativa (AERA, APA & NCME, 2014) recuperando tendencias en materia de evaluación educativa. El modelo para la certificación, incluye dos fases una de evaluación de conocimientos que se desarrolla mediante una prueba de selección de respuestas y una fase de evaluación de competencias. Cada fase incluye un curso de preparación previo a la evaluación, de manera que se puede considerar que se trata de un modelo de intervención-evaluación. En el presente artículo sólo abordaremos la evaluación de competencias. Es importante destacar que sólo las personas que superan la primera de evaluación pueden presentar la segunda fase que implica la evaluación de competencias.

Como inicio del trabajo de diseño y desarrollo del modelo de evaluación de competencias, retomamos la metodología de marcos de la buena enseñanza desarrollada en el ámbito educativo, para el desarrollo de los perfiles basados en estándares de competencia profesional de los principales actores que operan el sistema de justicia penal en México. Como parte de su elaboración se desarrolló un marco teórico en torno a las competencias implicadas en la práctica profesional de los operarios del sistema de Justicia Penal, la experiencia reportada en este artículo corresponde a la evaluación sólo de fiscales en virtud de que contamos ya con resultados de más de 800 personas evaluadas.

En el ámbito de la formación continua de personal docente en servicio, los marcos de la buena enseñanza, son un aspecto clave para promover el reconocimiento social de cualificación de las y los profesionales al proporcionar garantías de calidad de su desempeño (Meckes, 2007, 2013). Estos marcos son perfiles basados en estándares de competencia profesional, tendencia que se ha extendido a otros campos profesionales, y que parte del análisis colegiado entre especialistas, académicos y personal con experiencia laboral sobre las competencias que se requieren para el ejercicio eficaz de las funciones profesionales requeridas en un campo específico.

Su importancia no solo radica en que se constituyen en la base para el diseño y desarrollo de programas de formación y modelos de evaluación de desempeño profesional, lo cual contribuye a la mejora del ejercicio profesional (Manzi, González y Sun, 2011), sino en la generación de líneas de investigación en torno a los procesos cognitivos implicados en el desarrollo y adquisición de las competencias

representadas en estos marcos.

Con base en el perfil referencial y en el modelo cognitivo que lo sustenta, se procedió a la definición del dominio a evaluar para cada operario del Sistema de Justicia Penal en México. El trabajo implicó desarrollar un muestreo representativo de lo que debe saber y ser capaz de hacer una persona en un campo profesional determinado. Una tendencia actual que resulta muy útil para este propósito es la construcción de matrices de muestreo, herramienta de trabajo promovida por el National Research Council (NRC, 2012, 2013).

La construcción de estas matrices se llevó a cabo mediante la discusión colegiada de especialistas y académicos organizados en comités. La matriz se integró a partir de dos ejes, uno de los cuales refiere a todo el contenido teórico y práctico que requieren los fiscales, defensores, facilitadores y asesores jurídicos de víctimas para llevar a cabo las funciones que le corresponden a cada actor y el otro que refiere a las dimensiones de la práctica profesional representadas en el perfil basado en estándares de competencia. En las celdas de la matriz que se producen del cruce de las dimensiones de ambos ejes, se pudo especificar el desempeño esperado como resultado de este cruce. A partir de la definición del desempeño esperado fue más fácil construir tareas evaluativas relevantes que den cuenta del desempeño representado en las celdas de la matriz de muestreo.

Con relación a la contextualización de la evaluación como condición indispensable para evaluar constructos complejos, optamos por una evaluación auténtica basada en casos reales que los operarios del sistema seleccionan de su propia práctica para documentar las decisiones que tomaron, reflexionar sobre lo realizado e integrar evidencia de diferentes momentos del proceso penal, en un portafolios de evaluación. Con ello fue posible evaluar la capacidad de transferir conocimiento útil a una gran diversidad de situaciones y contextos en los que suceden los casos.

El desarrollo de las tareas evaluativas del portafolios de evaluación se llevó a cabo a partir de la definición del dominio a evaluar, mediante un proceso iterativo de desarrollo que incorporó la participación de un consejo técnico y diversos comités de expertos de las diversas regiones del país. Para ello retomamos conocimiento de teorías existentes que subyacen al concepto de competencias y aprendizaje profundo, así como a información relevante obtenida de estudios de investigación sobre criterios y procesos implicados en la toma de decisiones estratégica, la capacidad de argumentación y de comunicación oral y escrita, y la de solución de casos en situaciones reales.

También se utilizó información derivada de estudios de contraste entre profesionales egresados de programas de formación mediante currículo tradicional y currículo basado en problemas y estudios comparativos entre expertos y novatos, principalmente del campo de la medicina en donde existe una mayor tradición de investigación evaluativa (Leyva, 2011, 2018).

En gran medida el grado en el cual las tareas representan adecuadamente la conducta general de interés, refiere a la validez y es una cuestión importante acerca de si se han especificado adecuadamente los procesos cognitivos que subyacen a la tarea (Keehner, et al., 2013). Un modelo cognitivo proporciona una explicación fundada de los procesos cognitivos involucrados en los constructos objetivo de la evaluación, por lo que el modelo cognitivo está esencialmente ligado a las tareas o actividades que se diseñan para la obtención de evidencias directas del desempeño profesional.

Diseño del portafolios

El tipo de observaciones requeridas, orientó el diseño del portafolios con dos tipos de tareas para fiscales, aquellas más enfocadas a dar cuenta de las habilidades de dirección de la investigación y otras más enfocadas a valorar la ejecución en audiencias y juicios orales mediante videograbación de las

audiencias. En ambos tipos de tareas, el estímulo detonante fueron los casos reales con los que estos profesionales trabajan.

Tanto el formato de respuesta requerido en estos dos tipos de tareas, como los criterios de las rúbricas, permiten valorar habilidades transversales implicadas en el ejercicio de las principales funciones profesionales de fiscales dentro del sistema de justicia penal acusatorio, como es la atención digna y respetuosa a todos los intervinientes conforme a derechos humanos y la toma de decisiones estratégica.

Una condición necesaria por parte de las personas sustentantes para integrar el portafolios, fue la selección de casos diferentes para cada tarea. Se permitió el uso de un mismo caso para un máximo de dos tareas evaluativas. Una vez seleccionado un caso, la persona evaluada tenía que integrar toda la información requerida para cada tarea evaluativa, lo que implicó recuperar los documentos de su carpeta de investigación o de videograbaciones de audiencias, generadas durante el proceso de atención al caso, ello con el propósito de recuperar y reflexionar en torno a las acciones y decisiones que tuvo que realizar durante el proceso.

La integración del portafolios se llevó a cabo en línea mediante la plataforma Portafolios CEEAD, la cual se diseñó de acuerdo con las características de la evaluación para que en el transcurso de 10 semanas las personas sustentantes respondieran a las tareas evaluativas del portafolios y subieran los artefactos solicitados. El formato de respuesta, incluye espacios determinados en la plataforma para que las personas evaluadas respondan la tarea, y para que hagan la descripción de los hechos del caso seleccionado. La plataforma cuenta con un espacio para la incorporación de determinados artefactos –a manera de evidencias directas de su práctica- que nos permitieron verificar la pertinencia y veracidad de las respuestas, como son los videos de audiencias, los documentos, dictámenes periciales y diversos oficios que contiene la carpeta de investigación.

En el caso de las tareas escritas, se pide a la persona evaluada, que justifique algunas de las decisiones específicas que tomaron en los distintos momentos del proceso, de acuerdo con lo solicitado en cada tarea, lo cual representa un ejercicio de reconstrucción y síntesis de aquellos elementos que desde su propia perspectiva coadyuvaron de una u otra forma a la solución del caso bajo un enfoque de derechos humanos y en condiciones de igualdad para todos los intervinientes.

Por ejemplo, la recreación de los hechos, la recuperación de los elementos probatorios en los que sustentó su teoría del caso, la explicación de cómo utilizó la información disponible para definir sus líneas de investigación, o para sustentar alguna petición ante el órgano jurisdiccional correspondiente. En este tipo de tareas, los criterios de evaluación se centran en evaluar la congruencia y pertinencia de las acciones y decisiones tomadas, así como la capacidad de observación de las circunstancias de todos los intervinientes –víctimas, imputados, testigos- y en general al contexto específico en el que se desarrolló el caso.

En el caso de las tareas que evalúan ejecución en audiencias, el foco es la observación de la ejecución durante los procesos de sustentación de diversas peticiones ante el órgano jurisdiccional correspondiente, de acuerdo con las distintas etapas del proceso. En este tipo de tareas se solicita a la persona evaluada, igual que en la tarea escrita, que recupere la información relativa a un caso, y que una vez seleccionado el caso, recupere la videograbación de una audiencia o juicio oral, de acuerdo con lo que demanda cada tarea específica.

Una vez recuperada la videograbación, se le solicita que seleccione un fragmento de la videograbación en donde se muestre su ejecución de acuerdo con lo que especifica cada tarea. En cada tarea el foco se centra en evaluar la capacidad de argumentación en diferentes momentos del proceso de

sustentación de diversas peticiones como una vinculación a proceso o una solicitud de medidas cautelares; o en diversos momentos del proceso de litigación en juicios orales como un interrogatorio a un testigo o un alegato de clausura.

Integración del dictamen sobre el desempeño observado

En este tipo de evaluación no existen respuestas correctas o incorrectas, puede haber una gran variedad de respuestas para atender las características de los intervinientes ante una gran variedad de circunstancias y muchas de ellas pueden ser respuestas válidas, lo que representó un desafío técnico para la construcción de las rúbricas de calificación y para la integración de un dictamen global como resultado de la ejecución en diferentes momentos clave de la atención de los casos que suceden en gran diversidad de contextos.

Por ello la calificación de las respuestas requirió convocar a operarios pares de cada figura con un reconocido prestigio en el medio y a partir de la segunda aplicación a pares que obtuvieron la certificación bajo este modelo de evaluación. Todos ellos recibieron una capacitación general sobre el modelo de evaluación y se especializó a las personas convocadas en el uso de las rúbricas de no más de dos tareas del portafolios, mediante un entrenamiento con ejemplos y ejercicios prácticos en el uso de las rúbricas.

Como parte del proceso durante la capacitación y entrenamiento y a lo largo de toda la calificación se llevaron a cabo ejercicios de calibración para garantizar el apego a las rúbricas de calificación por parte del personal que califica. La calificación es doble ciego, es decir que las personas que califican no tienen acceso a la identidad de las personas evaluadas, ni las personas evaluadas saben cuáles personas califican sus respuestas y ejecución en las distintas tareas. Para ello se diseñó un algoritmo de asignación de respuestas a las personas que califican para evitar algún conflicto de intereses. Todo el proceso de calificación se lleva a cabo en 10 semanas.

Para la emisión del dictamen de si una persona tiene o no las competencias necesarias para el ejercicio profesional se recurrió a un modelo de calificación conjuntivo, es decir un modelo que incorpora más de una condición a diferencia de un modelo compensatorio en el cual sólo se considera la puntuación global (Martínez, Schweig y Goldschmidt, 2016). De esa manera se evitó la posibilidad de que alguien que alcanzó la puntuación global mínima, pero tuvo un desempeño por debajo de estándar de desempeño esperado en tareas esenciales para el ejercicio profesional, obtuviera la certificación. Una condición implica obtener una puntuación global en el portafolios igual o superior a un punto de corte y la otra condición es tener la puntuación mínima requerida en determinadas tareas evaluativas que son consideradas por los especialistas en el dominio a evaluar, como esenciales para el ejercicio profesional.

El modelo seleccionado implicó seleccionar una metodología para el establecimiento de los puntos de corte para diferenciar niveles de desempeño mediante una metodología basada en juicios de especialistas de la familia de los métodos holísticos. En general los métodos holísticos se utilizan principalmente en evaluaciones con fines de certificación de competencias profesionales, lo cual añade valor tanto a los participantes como a quienes otorgan la certificación porque se comunica lo que resulta importante para tener éxito en el área especial que se está examinando (Cizek & Bunch, 2007).

De los métodos holísticos existentes, se consideró que el más adecuado a las necesidades y el tipo de evaluación de Portafolios implementado para la evaluación de habilidades de los operarios del SJPA, es el denominado the body of work (BoW) desarrollado por Kingston, Kahl, Sweeney & Bay (2001), cuya principal característica consiste en la evaluación de conjuntos completos del trabajo de susten-

tantes por uno o más jueces que rinden un solo veredicto holístico o juicio global para cada muestra de trabajo.

La construcción de las rúbricas de calificación jugó un papel fundamental en la integración del dictamen y en la posibilidad de brindar una retroalimentación útil para las personas evaluadas y para las instituciones en las que laboran en materia de formación y profesionalización. Cada tarea evaluativa de los portafolios tiene diferentes criterios a valorar en cuatro niveles de desempeño. Los descriptores de desempeño se construyeron con base en el marco teórico y recuperan aspectos observables de las características de las respuestas propias de cada nivel.

En cuanto a los cuatro niveles de desempeño, se siguió un modelo teórico basado en estudios de investigación sobre diferencias entre personal experto y novel (Dreyfus y Dreyfus, 1980) así como en estudios sobre adquisición y desarrollo de competencias (Bogoya, 2000). De acuerdo con estos estudios existen diferencias cualitativas entre el desempeño en función del grado de consolidación de las competencias en determinado campo profesional.

Mientras que un desempeño experto implica la posibilidad de utilizar y transferir conocimiento útil a diferentes contextos y situaciones y presentar soluciones a partir del análisis integrado de todas las dimensiones de un problema, un nivel novel implica un desempeño que se basa en el análisis unidimensional o fragmentado de un caso, lo que implica algunas omisiones o carencias en las soluciones que ofrece, o que no siempre responde o atiende las principales características de los intervinientes o del contexto en el que suceden los casos.

Con base en estas diferencias cualitativas en nivel de desempeño, decidimos junto con comités de especialistas de cada figura del SJP, construir rúbricas analíticas y que en las rúbricas se colocaría en el nivel 3, el estándar de desempeño requerido o competente para el ejercicio de la función, en el nivel 4 un desempeño por encima del estándar, es decir un desempeño destacado propio de un profesional experto que ha logrado consolidar sus competencias para el ejercicio de sus funciones. En cuanto al nivel 2, los descriptores corresponden a un desempeño por debajo del estándar pero cercano al nivel de competencia requerido, es decir con algunas omisiones o debilidades que no afectan seriamente la solución del caso, mientras que en el nivel 1 se hace referencia a un desempeño lejano al estándar de desempeño esperado y que pone en riesgo la atención del caso.

A la fecha contamos con los resultados de cuatro aplicaciones de fiscales en 26 de las 32 entidades federativas del país. Los resultados entregados a cada persona evaluada, además del dictamen global, incluyeron una retroalimentación detallada por cada tarea evaluativa, con base en sus resultados en cada criterio de las rúbricas. Esto fue posible porque se construyó un algoritmo para la integración de los reportes, señalando sus fortalezas si en un criterio habían obtenido puntuación en el nivel de desempeño 3 (competente) o 4 (consolidado), o bien señalando un área de oportunidad en aquellos aspectos aún no logrados, si había obtenido nivel 2, o nivel 1. Las rúbricas de cada tarea tienen entre 3 y hasta 5 criterios de evaluación cada una, lo que permitió generar reportes con una gran riqueza en términos de la retroalimentación proporcionada.

Los resultados de estas aplicaciones, además nos permitieron llevar a cabo análisis psicométricos para obtener información de confiabilidad y validez. Para ello y acorde con las tendencias para este tipo de evaluaciones, se hicieron análisis de varianza con Teoría de la Generalizabilidad (TG), porque nos permite especificar y estimar los efectos de las facetas que intervienen en el modelo de evaluación y sus interacciones (Solano, 2017). Se hicieron análisis por tarea considerando las facetas: sustentante, criterio, codificador; y por portafolios completo, considerando las facetas: sustentante, tarea, codificador.

La importancia de haber realizado estos análisis, es que nos permitieron obtener la magnitud del

efecto de las tareas evaluativas y de los codificadores en los resultados obtenidos. En términos de su interpretación, lo que se espera de una buena evaluación, son valores de varianza mayores debido al desempeño de las personas evaluadas y valores significativamente menores en las otras dos facetas. Los resultados del análisis con TG obtenidos a partir de estos primeros procesos de evaluación, en general fueron aceptables y nos permitieron tomar decisiones para mejorar tanto los instrumentos como los procesos de calificación entre cada aplicación, lo que constituye un proceso de mejora continua del modelo de evaluación.

La validez de los argumentos de evaluación

En consonancia con los estándares internacionales en materia de evaluación (APA, AERA y NCEM, 2014) para obtener evidencia empírica es necesario contar con resultados de las evaluaciones y también llevar a cabo algunos de estudios de investigación para validar los principales argumentos del modelo de evaluación relacionados con las decisiones que se toman tanto a nivel individual como institucional de acuerdo con los usos previstos de certificación y profesionalización del personal que opera el Sistema de Justicia Penal en nuestro país.

Se requiere obtener evidencia del impacto de la evaluación en la práctica profesional de las personas que han participado en el programa, tanto los que han obtenido la certificación como las que no han podido concluirlo. Al respecto hemos iniciado un estudio sobre la percepción del impacto que ha tenido el modelo de evaluación en la práctica cotidiana de las personas certificadas.

A la fecha hemos recabado testimonios solo de algunas entrevistas a fiscales que han logrado la certificación y de algunas autoridades responsables de supervisarlos. Los resultados son aún incipientes, pero apuntan a validar los principales argumentos. A continuación, presentamos algunos de los hallazgos que hasta el momento hemos recabado. Las entrevistas se orientan a recuperar la percepción de las personas respecto a diferentes momentos y aspectos relativos a la experiencia de la evaluación, así como del desempeño profesional de las personas evaluadas. Como aún no se concluye el estudio, por ahora solo se reportan coincidencias en las percepciones relativas al tipo de impacto experimentado y cuando la coincidencia en alguna percepción proviene de personas con diferente función dentro de la organización. En este caso, se trata de triangular la información que proporcionan los vicefiscales con aquella que manifiestan las personas certificadas. Cuando la coincidencia en una percepción, es producto de la triangulación adquiere mayor credibilidad.

La opinión de las personas certificadas corrobora que:

- Las tareas evaluativas son muy cercanas a lo que significa su trabajo en el marco de la reforma al sistema de justicia penal (SJP).
- La integración de su portafolios les permitió reflexionar sobre la manera en que preparan sus audiencias, lo que ha redundado en una mejora en la organización de su trabajo y la manera en que estructuran y preparan sus argumentos.
- La experiencia tuvo una influencia positiva en sus creencias de autoeficacia, brindándoles más seguridad para realizar su trabajo cotidiano.
- Ahora tienen mayor conciencia del impacto que tiene su trabajo en las vidas de las personas intervinientes en los casos que desarrollan.
- Con respecto a la retroalimentación proporcionada en sus reportes individuales de resultados, si bien refieren fue útil para confirmar sus habilidades y para detectar algunas áreas de oportunidad, sugieren que se brinde información sobre acciones para mejorar su desempeño.

- Las tareas evalúan las habilidades asociadas a atender las circunstancias del contexto que son relevantes para la atención de los casos, mediante soluciones acordes a las condiciones del contexto y a las características de los intervinientes. La confirmación se basa en que las personas entrevistadas consideraron que a partir de las demandas de las tareas evaluativas del portafolios, pudieron dar cuenta del contexto de su práctica al incluir información específica en cómo se concretó el enfoque diferencial especializado, perspectiva de género, la atención a poblaciones vulnerables y a víctimas.
- Finalmente perciben un impacto en la forma en la que realizan su trabajo, de manera concreta mencionaron que ahora cuidan más la atención a la víctima, la clasificación preliminar del delito, así como lo que tienen que incorporar a la carpeta de investigación y las pruebas que utilizarán para acreditar su teoría del caso. Todo ello se puede considerar un impacto deseado del programa de certificación.
- Un aspecto emergente que surgió en las entrevistas es el hecho de que valoran el potencial que tiene la socialización de su experiencia al interior de la fiscalía, para mejorar el servicio a nivel institucional.

Si bien estos resultados coinciden con las tendencias observadas en un estudio de seguimiento que se llevó a cabo mediante una encuesta aplicada a nivel nacional, se tienen que tomar con reserva debido a que todavía son pocos las entrevistas realizadas con personas de la región norte del país. Lo importante a destacar es que la experiencia ha generado información muy valiosa que puede ser aprovechada, por lo cual concluimos este artículo con algunas lecciones aprendidas.

Algunas lecciones aprendidas

A nivel político, la aceptación del modelo de evaluación entre las instituciones de procuración de justicia y entre los operarios del sistema de justicia penal a nivel nacional, ha sido posible gracias al trabajo decidido y continuo de comunicación, sensibilización y difusión que se ha desarrollado con los principales actores implicados, logrando acuerdos y suma de voluntades en un proceso participativo a nivel nacional, logrando convenios y acuerdos indispensables para el desarrollo del proyecto.

Desde la elección del modelo de evaluación, se han conformado diversos cuerpos colegiados en los que participan especialistas, académicos, operarios y otros actores clave de las distintas regiones del país, que representan los principales intereses a nivel nacional. Todos ellos han aportado desde sus diferentes visiones, información importante para el diseño y desarrollo de los instrumentos desarrollados. Esta participación amplia y representativa ha jugado un papel decisivo en la credibilidad del programa. También el carácter formativo, el hecho de que la evaluación es voluntaria y que no tiene consecuencias negativas para quienes deciden evaluarse, han contribuido a la aceptación del programa entre las personas y las instituciones de justicia.

Otro factor clave para la implementación del programa fue que el CEEAD incorporó como parte de su staff a un equipo de profesionales expertos en evaluación conformado por psicólogos, pedagogos y psicopedagogos con amplia experiencia en evaluación educativa, para el diseño y operación de la evaluación que sustenta la certificación. El equipo ha sido el responsable de la conducción de los comités de especialistas para la construcción de los perfiles referenciales, la definición del dominio a evaluar, la construcción de las tareas y rúbricas, así como todas las decisiones sobre el modelo de calificación y el establecimiento de los puntos de corte.

El equipo es responsable también de la creación de documentos metodológicos y herramientas de trabajo para la capacitación de los especialistas que participaron en los distintos comités y en la cali-

ficación. El equipo es responsable de documentar todos los procesos implicados en el diseño e implementación de la evaluación, del análisis de la información derivada de las evaluaciones y la generación de reportes de resultados individuales e institucionales para soportar la toma de decisiones de las propias personas evaluadas y las decisiones a nivel de las instituciones de justicia. También el equipo tiene las competencias necesarias para llevar a cabo los estudios de investigación planteados en la agenda de validez.

A nivel técnico, la evaluación basada en competencias mediante la solución de casos en situaciones reales, reporta muchos beneficios no obstante las desventajas de orden práctico que hemos expuesto y que deben ser controladas, sobre todo en evaluación a gran escala. Los beneficios están asociados a la autenticidad y libertad que tienen las personas evaluadas para elegir los casos, reconstruir sus experiencias mediante la documentación en torno a las decisiones tomadas y los argumentos contruidos en una gran diversidad de situaciones que suceden en su entorno, así como a la observación de su propia ejecución en los videos de las audiencias y los juicios orales en los que han participado, todo lo cual contribuye a que puedan hacer un ejercicio de reflexión al contrastar su desempeño contra estándares expresados en un perfil basado en competencias.

La experiencia de integrar el portafolios, en sí misma tiene un valor formativo importante, como lo reflejan los primeros hallazgos obtenidos de los estudios de validez, lo que se complementa con la emisión de los reportes individuales de sus resultados con mensajes de retroalimentación a nivel de cada criterio evaluado para cada tarea evaluativa.

A nivel de investigación evaluativa, el desarrollo de constructos complejos es una tarea poco desarrollada, a pesar de su importancia para sustentar la validez de las evaluaciones. Algunas líneas de investigación recientes en el contexto psicométrico, se orientan a entender la importancia de la dimensionalidad de un constructo complejo para diseñar evaluaciones válidas (Gorin & Embretson, 2012). Por ello, el trabajo de validación debe enfatizar estudios orientados a la verificación de la validez de las interpretaciones de las medidas, más allá de demostrar su relación con otras medidas. Con el tiempo, hemos aprendido que sin una buena base teórica para predecir qué correlaciones deben obtenerse, carece de sentido hacer estudios basados en correlaciones entre medidas de diversos instrumentos.

Toda evaluación a gran escala, independientemente del método empleado, tiene un impacto en la enseñanza y en el aprendizaje. La naturaleza de este impacto no necesariamente es predecible, por lo que resulta deseable hacer estudios cuidadosos de posibles efectos intencionados o no. Las consecuencias de la interpretación o uso de una prueba deben ser evaluadas como un componente integral en los estudios de validación (Messick, 1995).

En términos prácticos, los resultados de estas líneas de investigación, resultan especialmente importantes para que los procesos de desarrollo y validación de evaluaciones de constructos complejos, se centren en la validación de la estructura latente, a partir de un marco conceptual que permita construir argumentos teóricos que después se puedan verificar empíricamente.

Finalmente, la factibilidad es un aspecto clave cuando se quieren implementar programas de evaluación y profesionalización docente que realmente contribuyan a mejorar la docencia. Es necesario destinar recursos a los programas de evaluación docente, tanto como a los de formación para su profesionalización. De acuerdo con la experiencia aquí presentada, un recurso muy importante es el capital humano experto en evaluación preparado para coordinar los procesos técnicos implicados en el diseño y desarrollo de los modelos de evaluación docente más apropiados. Al respecto, conviene involucrar al personal que se va a evaluar, desde la fase de planeación de cualquier modelo que se piense implementar, su participación es clave en todas las fases de construcción de un modelo que pretenda ser útil para

la mejora.

De manera específica para el programa, con base en los análisis psicométricos hasta ahora obtenidos, nos dimos cuenta que resulta necesario programar estudios de investigación para verificar el efecto que tienen los casos seleccionados por las personas evaluadas en los resultados de las evaluaciones. La información que se obtenga de estos estudios, nos permitirá afinar el modelo de evaluación y aportará información valiosa sobre las estrategias y procesos cognitivos implicados en la solución de las tareas evaluativas de los portafolios.

REFERENCIAS BIBLIOGRÁFICAS

- AERA, APA & NCME (2014), *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Barberá, E. (2005): La evaluación de competencias complejas: la práctica del portafolio. *Educere*, vol. 9, núm. 31, octubre-diciembre, 2005, pp. 497-504 Universidad de los Andes Mérida, Venezuela.
- Bogoya, D. (2000). *Una prueba de evaluación de competencias académicas como proyecto*. En: Bogoya, D. y colaboradores. *Competencias y proyecto pedagógico*. Santafé de Bogotá, D. C: Unibiblos.
- Busemeyer, J. R., & Diederich, A. (2010). *Cognitive modeling*. Los Angeles, CA: Sage.
- CEEAD (2019): Criterios y directrices de evaluación. En: https://cejume.mx/wp/wp-content/uploads/2020/09/200727_Criterios_evaluacio%CC%81n.pdf
- Cizek, G.J. y M.B. Bunch (2007), *Standard Setting: A guide to establishing and evaluating performance standards on tests*, California, SAGE Publications.
- Cook, D. A., Brydges, R., Ginsburg, S. & Hatala R. (2015) A contemporary approach to validity arguments: a practical guide to Kane's framework. *John Wiley & Sons Ltd. Medical Education*; 49: 560–575.
- Darling-Hammond, L., Falk, B. F. and Ancess, J. (2017): *Authentic Assessment in Action: Studies of Schools and Students at Work*, Teachers College Press.
- Dreyfus S. y H. Dreyfus (1980). A five-stage model of the mental activities involved in directed skill acquisition. *Operations Research Center, University of California, Berkeley*.
- Ercikan, K., & Oliveri, M. E. (2016). In search of validity evidence in support of the interpretation and use of assessments of complex constructs: Discussion of research on assessing 21st century skills. *Applied Measurement in Education*, 29(4), 310-318.
- Jornet, J. M.; González Such, J.; Suárez, J. M. y Perales, M. J. (2011). Diseño de procesos de evaluación de competencias: consideraciones acerca de los estándares en el dominio de las competencias, *Bordón* 63 (1), 2011, 125-145 • 143
- Kane MT. Validation. In: Brennan RL, ed. *Educational Measurement*, 4th edn. Westport, CT: Praeger 2006;17– 64.
- Keehner, M., & Smith, L. (2013). *Connecting actions, cognitions, and measurement: The role of cognitive science in NAEP TEL task development*. Paper presented at the Annual Meeting of the National Council on Measurement in Education. San Francisco, CA.
- Kingston, N.M. Kahl, S. R., Sweeney K. & Bay L. (2001) Setting performance standards using the Body of Work method. In Cizek, G. J. (Ed.), *Standard setting: Concepts, methods, and perspectives* (pp. 219-248). Mahwah, NJ: Erlbaum.

- Leyva, Y. (2011), *Validez de constructo en la evaluación de competencias médicas mediante pruebas referidas a criterio*, Editorial Académica Española.
- Leyva, Y. (2018): Una propuesta para el desarrollo profesional de docentes en México. Red. *Revista de evaluación para docentes y directivos*. Año 4. INEE. <https://www.inee.edu.mx/wp-content/uploads/2019/01/Red11.pdf>
- Manzi, J., González, R. & Sun, Y. (2011). *La evaluación docente en Chile*. MIDE UC: Centro de Medición.
- Markman, A. B. (1999). Knowledge representation. Mahwah: NJ: Lawrence Erlbaum Associates.
- Martínez, J. F.; Schweig, J. & Goldschmidt, P. (2016): Approaches for Combining Multiple Measures of Teacher Performance: Reliability, Validity, and Implications for Evaluation Policy. *Educational Evaluation and Policy Analysis*, Vol. 38, No. 4, pp. 738–756
- McEachen, J. (2017). Assessment for Deep Learning. Ontario, Canada: Fullan, M., McEachen, J., & Quinn, J. New Pedagogies for Deep Learning: A Global Partnership Retrieved: <http://teachingliterature.pbworks.com/w/file/attach/135692085/Assessment-for-Deep-Learning.pdf>
- Meckes, L. (2007). Evaluación y estándares: logros y desafíos para incrementar el impacto en la calidad educativa. Pensamiento Educativo, *Revista De Investigación Latinoamericana (PEL)*, 40(1), 351-371. Recuperado a partir de <http://pensamientoeducativo.uc.cl/index.php/pel/article/view/25519>
- Meckes, L. (2013). *Estándares y formación docente inicial. Profesores para una educación para todos. Proyecto estratégico regional sobre docentes* [Ponencia.]CREALC/UNESCO. III Reunión Técnica Estrategia Regional sobre Docentes. Santo Domingo, República Dominicana.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749. <https://doi.org/10.1037/0003-066X.50.9.741>
- Mehrens, W. (1992): Using Performance Assessment for Accountability Purposes. *Educational Measurement, Issues and Practice*. Vol. 11, NCME. <https://doi.org/10.1111/j.1745-3992.1992.tb00220.x>
- National Research Council (2012). *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/13398>.
- Pellegrino, J. W., & Hilton, M. L. (2012). *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century* (p. 257). Wash. DC Retrieved [Httpdownload Nap Educ. Cgi](http://download.nap.edu/cgi).
- Pellegrino, J. W., Wilson, M. R., Koenig, J. A., & Beatty, A. S. (2014). *Developing Assessments for the Next Generation Science Standards*. National Academies Press. 500 Fifth Street NW, Washington, DC 20001.
- Solano-Flores, G. (2017). Chapter 46: Generalizability. In D. Wyse, N. Selwyn, E. Smith, & L. E. Suter, (Eds.), *Handbook of Educational Research* (pp. 937-956). London, UK, Sage.