

Diseño de instrumentos para recoger datos en la evaluación de candidatos a ingresar en la DGI y de un modelo de análisis de resultados

Alberto Picón¹; Varinia Ferrando²; Juan José Goyeneche³

RESUMEN

Una actividad de relaciones con el medio aceleró el proceso de innovación en los instrumentos para recoger datos en pruebas de opción múltiple (POM), y con eso se mejoró la calidad del apoyo a las Unidades Curriculares de la FCEA que aplican este tipo de evaluación en las revisiones y exámenes. La herramienta se mantiene en un proceso de mejora continua y tiene en cuenta las necesidades de cada usuario.

En 2015 la Dirección General Impositiva hizo una convocatoria pública para cubrir 68 vacantes en sus cuadros administrativos. Se presentaron más de 12000 candidatos. Las autoridades decidieron que la selección inicial debía permitir a cada uno demostrar su capacidad para competir y, al mismo tiempo, era consciente de la imposibilidad de valorar todas las pruebas una a una personalmente. En otras oportunidades se había realizado una primera selección mediante sorteo, para aplicar las pruebas al número de candidatos que podían evaluar efectivamente. La DGI descartó este mecanismo y, en cambio, optó por una prueba de selección múltiple en la cual la UAE y el IESTA de la FCEA colaboraron mediante el diseño y procesamiento de los instrumentos para recoger datos en las evaluaciones iniciales; y posteriormente, el IESTA diseñó el modelo de análisis de los resultados. Los especialistas de la FCEA recogieron los lineamientos propuestos por el Tribunal del Concurso del Llamado 0005/2015 Administrativo Tributario - MEF - Dirección General Impositiva. Las mediciones incluyeron tanto la prueba de conocimiento como las pruebas psicolaborales de quienes superaron el nivel mínimo de la primera. El modelo para ambas pruebas fue diseñado y aplicado en el IESTA.

La experiencia permitió avanzar en el uso de instrumentos POM y en particular, en la capacidad de la UAE para satisfacer los requerimientos de la evaluación de aprendizajes en instancias con alto número de participantes. Además, se mejoró la competencia para atender las necesidades de datos de cada curso.

Palabras clave: evaluación de competencias, pruebas de opción múltiple, extensión y relaciones con el medio, modelos estadísticos, selección de personal.

La FCEA tiene una dilatada trayectoria en extensión y relaciones con el medio, generalmente a través de sus institutos, de estadística y de economía. Sin embargo, también los Departamentos de Contabilidad y de Administración han participado en proyectos sociales y productivos. Desde la vigencia de la nueva ordenanza de grado (UDELAR, 2011) y su consecuente reflejo en los planes de estudio de la facultad (UDELAR, 2012), estas actividades no solo son permanentes, sino que forman parte de los trayectos de acreditación de las carreras. Existe una Red temática de Emprendedurismo de la UDELAR (EMPENUR), cuya sede y coordinación está en la FCEA. Hay una Unidad de Extensión y Relaciones con el Medio (UERM) que coordina los Espacios de Formación Integral (EFI), en los cuales se realizan las actividades de extensión creditizada. Otra unidad central es la Unidad de Apoyo a la Enseñanza (UAE), que desde 2014 está equipada y cuenta con los recursos para apoyar el desarrollo y la realización de pruebas de evaluación de aprendizajes que incluyen preguntas de opción múltiple (POM). Esta capacidad le permitió colaborar con la Dirección General Impositiva (DGI), en el proceso de selección de 68 nuevos funcionarios; y con esta actividad de relaciones con el medio se sumó a las demás que hasta ahora las han venido desarrollando. Además, dado que estos procedimientos son comunes en la

¹ UDELAR-Facultad de Ciencias Económicas y de Administración. Unidad de Apoyo a la Enseñanza.

² MEF-Dirección General Impositiva. Grupo de Trabajo Concursos Internos.

³ UDELAR-Facultad de Ciencias Económicas y de Administración. Instituto de Estadística.

Administración, no se trata de una experiencia aislada; otras instituciones han recurrido para asesorarse.

Presentación del caso

En 2015 la DGI convocó a un concurso abierto para llenar 68 vacantes de Administrativo Tributario (DGI, 2015). Se inscribieron más de 12000 personas. La DGI se puso en contacto con la FCEA para realizar una primera selección mediante una prueba con POM. La situación inicial era aplicar una prueba a los 12380 candidatos, basada en conocimientos generales y nivel de estudios mínimos equivalentes a secundaria. Se presentaron efectivamente 5000, de los cuales alrededor de 1000 superaron el mínimo. A estos candidatos se les aplicaron dos pruebas de evaluación psicolaboral. La prueba de conocimiento contenía 25 ítems con cuatro opciones. La primera parte de la psicolaboral contenía 132, con respuestas mediante una escala de Likert de 5 a 1. El segundo test era una serie de 50 ítems con arreglos de piezas de dominó – test de dominó -, la respuesta indicaba los valores de la pieza faltante y su dirección, 6-0 no 0-6, 1-2 no 2-1, 6-6, y así sucesivamente.

En mayo de 2015 el Ministerio de Economía y Finanzas convocó un Llamado a concurso público y abierto para la contratación de 68 Administrativos Tributarios (para Montevideo y el Interior del país), en régimen de contrato de Provisorio, para desempeñar tareas en la Dirección General Impositiva.

El portal de Uruguay Concurra recibió vía electrónica la postulación de más de 15.000 candidatos, de los cuáles fueron preseleccionados 12.388 por el sistema informático del Área de Reclutamiento y Selección de la Oficina. La preselección se realizó controlando que lo declarado por los candidatos en el formulario de postulación concordara con los requisitos excluyentes de formación y experiencia exigidos en las Bases del llamado.

Al mismo tiempo, se constituyó el Tribunal del Concurso designado para establecer el cronograma y gestionar las actividades del proceso de selección. El Tribunal fue designado por el Jefe de la DGI y se conformó con tres miembros: un representante de la Dirección de la DGI, un representante de la Oficina Nacional del Servicio Civil (ONSC) y una persona idónea en la materia del concurso. El Tribunal contó también con la presencia de un veedor de la Confederación de Obreros y Funcionarios del Estado (COFE). Se aplicó el Manual de reclutamiento de la ONSC (ONSC, 2015).

El Tribunal estableció que los factores a evaluar en el Concurso se ajustarían a los siguientes porcentajes:

- 1) Prueba de Conocimiento: 50 puntos
- 2) Méritos y Antecedentes: 25 puntos
- 3) Evaluación Psicolaboral: 25 puntos

El candidato aprobaría el concurso obteniendo como mínimo 70 puntos de los 100 puntos totales, y sería eliminado del proceso de selección el candidato que no tuviera oportunidad matemática de alcanzar dichos puntos mínimos:

“aquel postulante que con el puntaje obtenido en las etapas realizadas, no le permita con el máximo de puntos restante, completar el total de 70 puntos mínimos exigidos para la aprobación del Concurso” (Artículo 30, Decreto 223/013).

Prueba de conocimiento

En los concursos de Oposición y Mérito, los candidatos compiten entre sí acreditando sus méritos y antecedentes, y realizando pruebas con el fin de determinar cuáles son las personas más adecuadas para el puesto de trabajo. Para determinar el grado de conocimientos y habilidades específicas poseídas por los concursantes, el Tribunal resolvió realizar una prueba de conocimiento sobre temas relacionados con las tareas a realizar y con el perfil requerido.

Dada la numerosidad de los sujetos a evaluar, y teniendo en cuenta los tiempos legales del proceso de selección y las urgencias organizacionales del llenado de vacantes, el Tribunal resolvió realizar un cuestionario e implementar un procedimiento de aplicación y corrección de prueba que le permitiese realizar la evaluación en la forma más objetiva, eficaz y rápida posible.

Para ello el Tribunal resolvió contactarse con la FCEA y de las reuniones mantenidas surgió la idea de aplicar una prueba con POM y proceder a su corrección a través del lector óptico. A partir de esa decisión se distinguen, sucesivamente, las siguientes actividades: i) diseño de la prueba; ii) cantidad de pruebas o versiones; iii) locaciones de aplicación; iv) aplicación.

¿Cómo se diseñó la prueba de conocimiento? La prueba de conocimiento versó sobre la bibliografía proporcionada y fue diseñada por el idóneo del tribunal. Elaboró un cuestionario de 25 preguntas con cuatro opciones cada una, de las cuales solo una era correcta. Cada pregunta valía dos puntos si se contestaba correctamente y cero si la respuesta era incorrecta o no se respondía. En ningún caso se restaron puntos.

¿Cuántas pruebas se realizaron? Se elaboraron 6 pruebas y a través de EDUCANTEL⁴ se realizó el mezclado de las 25 preguntas y 4 opciones, resultando así 24 versiones necesarias de pruebas para distribuir en los distintos centros de aplicación.

¿Cuántos centros de aplicación de prueba se organizaron para la toma? Las pruebas de conocimiento se realizaron los días 25 y 26 de julio de 2015. El día sábado 25 se realizaron las pruebas en forma simultánea en dos locales diferentes. En el Palacio “Cr. Gastón Güelfi”⁵ se programaron tres tomas de 2.000 personas cada una y en el Laboratorio Tecnológico del Uruguay (LATU)⁶ tres turnos de 1.000 personas cada una. El domingo 26 de julio las pruebas fueron realizadas en el local del LATU en tres turnos de 1.000 personas cada uno. Las locaciones preveían la concurrencia de aproximadamente **12.000** personas, hecho que no se concretó ya que efectivamente concurrieron **4.860** a rendir la prueba.

¿Cómo se organizó la toma de la prueba de conocimiento? Se formaron dos equipos de trabajo que colaboraron con el desarrollo de la prueba. Setenta personas fueron asignadas al Palacio “Cr. Gastón Güelfi” y aproximadamente otras 30 al LATU. Dentro de cada equipo se conformaron parejas responsables de controlar el ingreso de los concursantes, de entregar los cuestionarios y recepcionar los protocolos de respuestas. Los restantes funcionarios fueron distribuidos en los locales con el cometido de vigilar el normal desarrollo de la prueba.

⁴ Educantel es la comunidad educativa del Área de Relacionamiento con la Comunidad de la Administración Nacional de Telecomunicaciones (ANTEL). Disponible en <https://educanet.anel.com.uy/>. Visto 25set2016.

⁵ Tiene un aforo aproximado de 4800 personas. Disponible en <http://www.ultimasnoticias.com.uy/Edicion-UN/articulos/prints-20ene2011/dep04.html>. Visto 25set2016.

⁶ Cuenta con un Centro de Reuniones y Eventos (CRE) con salas que superan en conjunto un aforo de 3000 personas. Disponible en <http://latu21.latu.org.uy/parquetecnologico/index.php/cre>. Visto 25set2016.

Prueba Psicolaboral

Luego de la prueba de conocimientos, y de la instancia de evaluación de méritos y antecedentes, aquellos postulantes que cumplieron con los requisitos excluyentes y obtuvieron la valoración de 45 puntos o más pasaron a la instancia de la Prueba Psicolaboral.

De los 3.719 concursantes que presentaron la documentación acreditando su experiencia y su formación: **966** personas obtuvieron o superaron los 45 puntos; 1.896 personas no superaron los 45 puntos y 857 personas no cumplieron con los requisitos excluyentes o no presentaron documentación suficiente.

La evaluación psicolaboral consiste en aplicar distintas técnicas de evaluación, con la finalidad de identificar en los potenciales candidatos los conocimientos, habilidades y aptitudes requeridas para los puestos de trabajo. En la actualidad se acepta que las predicciones más ajustadas se obtienen utilizando una combinación de diferentes métodos, de los que se extrae información buscando visualizar las conductas de los sujetos y procurado predecir el comportamiento de los candidatos ante las distintas situaciones y funciones.

La evaluación laboral no persigue clasificar a los sujetos en categorías diagnósticas, sino que pretende encontrar al candidato cuya características psicológicas sean las más adecuadas para el puesto a ocupar.

El Tribunal del Concurso, considerando que el tiempo insumido para realizar la evaluación psicolaboral de 966 candidatos, con los recursos humanos disponibles en la organización, excedería los plazos establecidos en el cronograma del Llamado decidió contratar a una consultora externa para diseñar la batería de técnicas a utilizar para identificar y evaluar las competencias específicas definidas en el perfil de las Bases del Concurso. Paralelamente, en base a los buenos resultados obtenidos en la aplicación y corrección de la prueba de conocimiento, se decidió contratar nuevamente los servicios de la FCEA para agilizar la toma y la obtención de resultados de las prueba.

¿Qué técnicas se utilizaron realizar la evaluación psicolaboral?

Las competencias conductuales y/o funcionales establecidas en las Bases del llamado para el puesto fueron: Orientación al ciudadano, comportamiento ético con el servicio público, compromiso, proactividad, capacidad para aprender y comunicación efectiva.

Spencer y Spencer (1993) define las competencias como “característica subyacente de un individuo que está casualmente relacionado con un estándar de efectividad y/o con un desempeño superior en un trabajo o situación”.

Se podría decir que una competencia es una construcción a partir de una combinación de recursos internos (aptitudes, conocimientos adquiridos, rasgos, motivación, saber hacer) y recursos del contexto donde se inserta la organización, como el sistema sociocultural, los sistemas políticos a los que se somete, la cultura organizacional, etc. que se movilizan en conjunto para lograr el desempeño esperado. Según Martha Alles (2010) las competencias son características de personalidad, que devienen en comportamientos que generan un desempeño exitoso en un puesto de trabajo.

En ese marco, para identificar y evaluar dichas competencias, la consultora externa sugirió la aplicación de tres test: un inventario de personalidad (Big Five Questionnaire o BFQ), un test de nivel (Test de Inteligencia General o nivel 2 o TIG 2) y un test proyectivo (Test de Diseño Wartegg o Test de Wartegg). El BFQ y el TIG 2, además de permitir indagar rápidamente las competencias

requeridas, son de fácil aplicación y corrección, y por tanto muy adecuadas para aplicarse colectivamente.

El **BFQ** tiene como finalidad evaluar, mediante 132 elementos, 5 dimensiones de la personalidad: energía, afabilidad, tesón, estabilidad emocional y apertura mental. En cada una de estas dimensiones se identifican dos subdimensiones: energía y dominancia; cooperación-empatía y cordialidad-amabilidad; escrupulosidad y perseverancia; control emocional y control de impulsos, y apertura a la cultura y a la experiencia. Con la finalidad de adecuar el test a los requerimientos del concurso, cada dimensión y subdimensión se estableció según el comportamiento asociado planteado por la Oficina Nacional del Servicio Civil (ONSC) para cada competencia.

Adicionalmente, el test brinda una escala de distorsión que consta de 12 preguntas, que tiene como finalidad proporcionar una medida de la tendencia del sujeto a ofrecer un perfil falseado de sí mismo al responder al cuestionario.

El **TIG 2** es un test de inteligencia dirigido a evaluar la capacidad para conceptualizar y aplicar el razonamiento sistemático a la resolución de problemas. Permite evaluar el Factor “g”, entendido como la capacidad de abstracción y de realización de trabajo intelectual. En el concurso, su aplicación permitió evaluar la capacidad de aprendizaje. Esta prueba consta de 50 recuadros que representan un grupo de fichas de dominó, donde en cada grupo de fichas hay una que está en blanco y tiene que completarse en forma correcta.

Es de esperar que en los procesos de selección los sujetos tienden a distorsionar las respuestas, al intentar contestar lo que creen será considerada una buena respuesta por el evaluador, dando respuestas que no coinciden exactamente con su autovaloración honesta. Es por ello que para analizar la personalidad de los candidatos a seleccionar se entendió necesario la utilización de técnicas proyectivas.

En el ámbito laboral, el psicodiagnóstico no reside en una medición objetiva ni en una apreciación intuitiva, sino como Cristina Weigle (1996) señala partiría del registro de los observables, de las manifestaciones clínicas verbales y paraverbales que son expresadas tanto en entrevistas como en las técnicas proyectivas.

En las técnicas proyectivas el estímulo proyectivo moviliza ciertas respuestas a partir de las cuales pueden inferirse las fantasías inconscientes, que expresan la vida psíquica del sujeto. Estas fantasías inconscientes se inferirán del material proyectivo a través del proceso perceptivo que el sujeto realiza frente a la consigna y al estímulo.

Ante la imposibilidad material de realizar entrevistas individuales a 1.000 postulantes por los tiempos que eso insumiría, se optó por incorporar a la batería el test de Diseño Wartegg. Este test es una técnica proyectiva semiestructurada muy utilizada en el ámbito laboral ya que permite realizar una descripción global de la personalidad del sujeto, considerando los aspectos cognitivos, afectivos, interpersonales, de identidad y eventualmente posibilitan detectar elementos psicopatológicos.

Es un test gráfico de la forma que ofrece distintos signos y dibujos, que motivan a continuar con esos estímulos. Es una técnica rápida, simple y barata, de cómoda aplicación, ejecución e interpretación que nos permitió comprobar lo surgido por las otras dos pruebas psicolaborales.

La aplicación se realizó en una hoja original de prueba en la que constan 8 cuadros blancos insertos sobre un fondo negro, dentro de los cuales hay pequeñas señales. Se solicitó a los concursantes completar cada cuadro con un dibujo, pudiendo escoger el cuadro por el cual comience y el orden subsiguiente, dejando la anotación de la secuencia de los cuadros que dibuja. De esta forma, a

través de la producción de cada sujeto examinado pueden abordarse cuatro funciones básicas de la personalidad: emoción, imaginación, energía e intelecto. Se muestra en la figura 1.

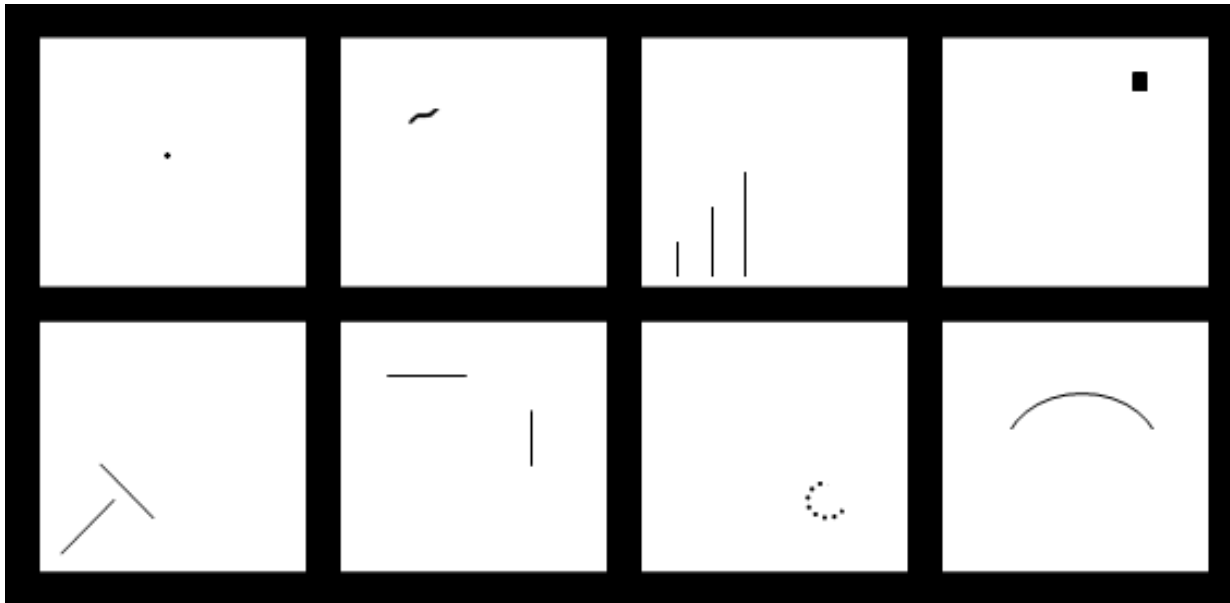


Figura 1. Test de Diseño Wartegg

¿Cómo se organizó la prueba psicolaboral? La prueba psicolaboral se realizó el día 7 de noviembre de 2015 en el LATU, previéndose aplicar a un total de 966 personas. La toma de los test estuvo a cargo de una Psicóloga del Departamento de Recursos Humanos de la DGI y se aplicaron a los 897 postulantes que se presentaron a rendir la prueba.

El test de Diseño Wartegg fue corregido por dos psicólogas de la DGI. Las otras dos pruebas, BFQ y TIG2 fueron corregidas a partir de los resultados recuperados por el lector óptico de la FCEA.

Extensión y relaciones con el medio

La extensión es una actividad universitaria que realizan los docentes. La UDELAR, y por lo tanto la FCEA, consideran actividades de extensión:

“...aquellas que impliquen colaboración entre actores universitarios y no universitarios en procesos interactivos donde cada actor aporta sus propios saberes y todos aprenden. Se trata de actividades que contribuyen a la creación cultural y al uso socialmente valioso del conocimiento, a la comprensión y a la solución de problemas de interés general y en especial en referencia a los sectores más postergados” (UDELAR-FCEA, 2013).

El asesoramiento y apoyo a la DGI, en las instancias correspondientes a las pruebas de selección del concurso para cubrir las 68 plazas, encuadra en la definición anterior. Es más, la experiencia de la UDELAR ha llevado a refinar la definición en el sentido de desagregar dos líneas para su implementación: la extensión y las relaciones con el medio.

Con el enfoque redefinido en la primera década de este siglo, muchas actividades que no se consideraban como extensión ahora lo son. Por ejemplo, cuando se ofrece actualización a los egresados o cualquier otra oferta educativa que no esté asociada a un título y a la que pueden acceder personas que no los posean, es una actividad de relaciones con el medio. Cuando

organizaciones de cualquier naturaleza reciben apoyo de la FCEA, en cualquier área de su interés o necesidad, estamos, otra vez, ante una actividad de este tipo.

La UAE, como unidad de apoyo de las actividades de enseñanza, está enfocada en los procesos de enseñanza y aprendizaje de las carreras de grado. Sin embargo, desde su establecimiento hasta el presente la oferta educativa creció, y con ella las demandas. Así, por ejemplo, en lo que puede considerarse un componente del control de calidad de los cursos, se realizan encuestas a los estudiantes en cursos de grado y posgrado. Y más recientemente se incursionó en la evaluación de aprendizajes. La UAE está en condiciones de diseñar instrumentos de colección de datos de las pruebas ajustadas a las necesidades de cada Unidad Curricular (UC). No obstante, la tecnología y los recursos con que cuenta tienen aplicación más allá de su ámbito de origen, y es razonable que se ponga a disposición de la sociedad en la medida de sus posibilidades.

El IESTA tiene una tradición de actividades de extensión desde su creación en 1942 “como ámbito de estudio, actualización y elaboración de trabajos en el campo de la Estadística”⁷; y donde “se desarrollan, además, actividades de extensión, formación, intercambio y difusión, en un proceso de retroalimentación permanente entre ellas”⁴. En el caso de la DGI, han tomado a su cargo el análisis de los resultados de las pruebas iniciales y, en particular, elaboró el modelo de análisis de las pruebas psicolaborales, a partir de las necesidades discutidas y de la relación entre las variables identificadas por el Grupo de Trabajo Concursos Internos (GTCI), de la DGI.

Es razonable afirmar que en estos momentos la FCEA puede desplegar todos sus recursos en actividades de extensión y relaciones con el medio, porque los mismos recursos con que se aborda la oferta educativa pueden ser aplicados con éxito para contribuir “a la creación cultural y al uso socialmente valioso del conocimiento, a la comprensión y a la solución de problemas de interés general”, como prevé la definición al uso.

Preguntas de opción múltiple (POM)

Puede rastrearse el origen de las preguntas con respuestas de opción múltiple en la historia de los censos. Los Estados Unidos de América (EUA) los han realizado desde 1790⁸. Y puede señalarse la obra *Educational Psychology* (Thorndike, 1903) como el catalizador que reunió los estudios sobre el desempeño de los alumnos en edad escolar y aquella técnica. Se considera que el trabajo de Frederick Kelly, en la selección de reclutas para las fuerzas armadas, el primer hito en el uso de los test de opción múltiple para medir la capacidad de las personas y seleccionarlas en orden a su preparación o formación posterior. Un hito en el desarrollo de la tecnología educativa y de la selección de personal. En el primer caso para medir conocimientos y predecir resultados futuros; en el segundo, para ubicar a cada empleado en el puesto adecuado a su capacidad. Y todo porque había una Gran Guerra que demandaba tropas y quien las liderara en el campo de batalla: soldados y oficiales.

A pesar de que el propósito de las POM es presentar a los participantes un método equivalente y objetivo para medir su desempeño en un área dada, es posible que la prueba contenga sesgos ideológicos a partir de las creencias del diseñador, por eso es conveniente, y actualmente se realizan de ese modo, que la propuesta sea el producto de un equipo de trabajo. Entre 1917 y 1918 alrededor de 1.7 millones de hombres fueron examinados mediante pruebas estandarizadas diseñadas para seleccionar futuros oficiales o soldados de línea en los EUA. Sin embargo, el test contenía defectos en su diseño, pues partía de una perspectiva racista y la prueba confirmaba las expectativas de los evaluadores blancos sobre el comportamiento de los afroamericanos (Watters, 2015).

⁷ <http://www.iesta.edu.uy/>

⁸ <http://www.census.gov/>

Es conocido que un evaluador puede sesgar la medición de diferentes aspectos a partir de los resultados en uno de ellos (Thorndike, 1920) y que se conoce como efecto de halo, efecto *Pigmalion* y otros nombres bastante populares; así como que dos docentes corrigiendo un mismo grupo de pruebas obtiene resultados diferentes, ya sea porque aplican otra perspectiva o por favoritismo (Kamenetz, 2015, pág. 50). Mientras que la diferencia de criterio puede reducirse mediante pautas de corrección que honestamente aplica el equipo de corrección, el sesgo es subjetivo. Hay quienes abogan por la eliminación del docente en el proceso de evaluación, mediante la aplicación de POM. Según el pensamiento de Kelly (1916), enseñanza efectiva significaba resultados uniformes, un pensamiento influido grandemente por la línea de producción promovida por Ford: “apreciando un producto confiable, uniforme y fácilmente replicado (...) por encima del ingenio, la creatividad, la individualidad, la idiosincrasia, el juicio crítico, y la variabilidad” (Davidson, 2011a, pág. 115ss). Según Davidson, Kelly no pretendía que las pruebas estandarizadas se expandieran como lo hicieron. En realidad este tipo de prueba tenía como objetivo medir el pensamiento de orden inferior (“lower order thinking”) en las masas. Las POM “se han convertido en el patrón oro” en los EUA desde las pruebas escolares hasta las que se realizan en escuelas profesionales de posgrado (Davidson, 2011b). Es fácil de aplicar y corregir, especialmente indicado donde hay gran cantidad de personas y su carácter estándar expone a todos a las mismas alternativas.

Luces y sombras de las POM

Las POM son una herramienta válida y confiable para medir competencias en contextos determinados. Es decir, no sirven para todo. Es oportuno utilizar POM para medir el aprovechamiento de un estudiante, o el nivel de capacitación, conocimientos generales, en fin, regularidades que se espera encontrar en la mayoría de los sujetos sometidos a la prueba. Es inútil para medir talentos inesperados.

Cuando pensamos en un grupo de personas con un perfil concreto, específico, éste se define por anticipado y se diseña una batería de preguntas que directa e indirectamente den cuenta del grado en que dicho perfil está presente en cada sujeto evaluado. Es el caso de los conocimientos y competencias profesionales que se desarrollan durante el proceso de formación. Por eso, cuando un individuo obtiene su título profesional se espera que sea capaz de abordar los temas afines a su carrera, con la misma solvencia, o al menos en un entorno de competencia, aceptado tanto por el colectivo científico como por la sociedad. La sociedad cuenta con la garantía de un producto satisfactorio, el servicio profesional de que se trate, sin importar cuál de ellos lo realice. La evaluación durante la formación debe ser necesariamente idéntica para todos, es decir estándar. Este es un espacio apropiado para aplicar POM.

La numerosidad es el otro ingrediente relevante para tomar decisiones sobre POM. Especialmente cuando se trata de clasificar un número importante de individuos y sobre todo, para seleccionar, como en el caso de la DGI: más de 12000 inscriptos. Es evidente que corregir pruebas es una actividad agotadora. El desgaste que experimenta un corrector solamente podría atenuarse con un régimen de trabajo que, o bien, extendería el período de corrección más allá de lo aceptable, o bien, demandaría un número de correctores entre los cuales la pauta podría desvirtuarse, aun sin que mediara la voluntad de infringirla. Los errores podrían disminuir con una corrección recurrente, otro u otros correctores y revisores, pero esto solo profundiza el problema del tiempo. Una POM parece la solución: el trabajo extenso ocurre durante el diseño previo a su aplicación, que por otra parte, afectará a todos los sujetos por igual.

El lado oscuro de las POM es que hace falta tener claro qué se quiere medir y asegurarse de que el instrumento, o mejor dicho, el conjunto de ítems que lo componen, den cuenta efectivamente del

estado y comportamiento que se pretende observar. A mediados del siglo XX un equipo de investigadores en los EUA identificó una serie de objetivos de la educación en todos los niveles, que han sido la base de trabajos de investigación y de aplicación de pruebas desde entonces. Se conocen como taxonomía de Bloom (Bloom, 1956). Cuando se trata de seleccionar personal para un puesto de trabajo hay dos tipos de conocimientos: i) los del nivel de estudios mínimos que se requieren para ocuparlo; ii) los generales que se han definido en el perfil según el imaginario objetivo o subjetivo del equipo evaluador.

Lo evidente y lo implícito: lenguaje superficial y lenguaje profundo de las POM

Una evaluación basada en POM incluye una cantidad de ítems con una estructura esencialmente similar. Un enunciado que se completa con dos o más opciones alternativas que pueden o no ser excluyentes. El caso mínimo es una propuesta dicotómica con opciones: verdadero y falso. Los otros incluyen de tres a cinco, una es la respuesta las demás son distractores. Los distractores deben tener apariencia de ser respuestas posibles, de lo contrario no cumplen su función, como se muestra en la figura 2 (Picón Martínez, 2016). La prueba general debe estar construida de manera que pueda descartarse el azar para alcanzar el nivel mínimo de suficiencia.

Ítem bien construido	Ítem mal construido
El altímetro indica la distancia vertical desde el instrumento hasta el a. piso. b. nivel de presión 1013.2 hPa. c. nivel medio del mar.	El altímetro indica la distancia vertical desde el instrumento hasta el a. nivel de la presión en la superficie. b. nivel de presión 1013.2 hPa. c. nivel de la presión a nivel del mar.

Figura 2. Ejemplo de POM

El ejemplo de ítem mal construido contiene los dos errores más evidentes: i) texto innecesario que ocupa tiempo excesivo sin aportar evidencia para el evaluador, y ii) la posibilidad de deducir la respuesta. Efectivamente, en este último caso el estudiante que tiene el conocimiento básico de que el altímetro funciona midiendo la presión circundante puede deducir que un instrumento instalado en una aeronave no puede medir el valor de la presión atmosférica más allá de su entorno y por lo tanto descartará las opciones a. y c., pero no puede concluirse que conoce las implicancias de la respuesta, opción b. como sí se puede concluir del ítem correcto. Como se ve en ese ejemplo, todo el texto que puede aplicarse a las tres opciones se incluye en el enunciado, lo cual sugiere que todos los ítems deberían construirse reduciendo la extensión de las opciones a las mínimas palabras posible, preferentemente una (Rodríguez Ayán, 2014, págs. 3-4).

El azar se puede amortiguar mediante una distribución equilibrada de respuestas; para cuatro opciones, como se utilizaron en el caso de la DGI, sería conveniente que haya un cuarto de respuestas en cada una: un cuarto a., un cuarto b., un cuarto c., un cuarto d. Algunas pruebas que ponen el nivel de suficiencia igual o superior al 50% de respuestas correctas descartan la cantidad de respuestas correctas que equivalen a un cuarto del total de POM, de manera que la aprobación requiere un 62,5% de respuestas correctas como mínimo. Existen muchos otros procedimientos para medir el resultado de una prueba con POM y contrarrestar el azar, por ejemplo, penalizando los errores o aportando puntos por no responder al azar, es decir, por dejar el ítem “en blanco” cuando no se tiene certeza de la respuesta.

Estos elementos más o menos evidentes pueden considerarse el lenguaje superficial de las POM, pero: qué significan las respuestas, conduce a interpretar el lenguaje profundo (Picón Martínez, 2016). En este sentido, discutiremos la asignación de puntos o determinación del nivel de suficiencia, el significado de algunos errores, y la evidencia combinada de POM.

En general las pruebas con POM se miden linealmente: todas los ítems tiene el mismo valor en relación a la calificación y un número n de respuestas correctas asegura el nivel de suficiencia. Se pueden hacer ajustes por errores, bonificaciones por evitar el azar o contabilizaciones como el caso de 62,5% mencionada más arriba. Cuando reflexionamos sobre esta asignación lineal y la relacionamos con las destrezas de nivel inferior y superior de la taxonomía de Bloom, nos damos cuenta de que en los hechos es poco probable que todas las POM tengan el mismo nivel de dificultad. Por lo tanto, que se les asigne el mismo valor no da verdadera cuenta del comportamiento que se pretende medir. Especialmente porque los recursos y la intensidad con que los aplica el evaluado en cada caso son diferentes. No es lo mismo tener memoria que tener criterio para evaluar una situación y aportar una respuesta que requiere considerar varios datos.

Existen pruebas psicotécnicas en las que se aplican POM. Estas evaluaciones tienen en cuenta la combinación de respuestas a varios ítems o una misma POM es utilizada en combinaciones diferentes para anticipar comportamientos en diferentes áreas. Este tipo de POM no tiene, por así decirlo, respuestas correctas o incorrectas; y en ese sentido, no satisfacen los criterios de POM para medir el resultado de un curso curricular, por ejemplo.

Una prueba modélica sea con o sin POM debería aportar datos que permitan medir realmente lo que se pretenda observar. Tales pruebas están sujetas a las mismas restricciones o requisitos de validez y confiabilidad de cualquier variable utilizada en un proyecto de investigación.

De los modelos estadísticos y su aplicación a la selección de personal

Los test son estímulos a un comportamiento, en una situación homogénea o estandarizada, para todos los sujetos. El comportamiento se evalúa a partir de una comparación estadística con el observado en un número de otros sujetos expuestos a la misma situación. Los diferentes resultados dan lugar a clasificaciones y tipologías que responden al marco teórico de todo el experimento o prueba.

Los test psicológicos, sea de competencia o de actitudes, generalmente proceden de contextos donde el número de sujetos disponibles permite refinar instrumentos – los test – que producen datos y la consecuente información de calidad. En esas condiciones, no es extraño que muchos se hagan populares fuera de los contextos de origen. La pregunta que se hace el investigador es si los resultados en el nuevo entorno mantienen la validez y confiabilidad de los obtenidos en el origen. Esta duda razonable y científicamente contrastable da lugar a numerosos proyectos en los que se intenta constatar precisamente eso: la aplicabilidad del instrumento o, en general, la pertinencia de una teoría para explicar fenómenos que no estaban en el terreno donde fue elaborada. Los test de comportamiento actual o esperado - competencias y actitudes -, contrastan a cada sujeto con la media estadística de una población o universo, cuyo valor, a su vez, procede del análisis de una muestra. El comportamiento promedio o norma es influido por muchas variables, entre ellas el contexto, la cultura del grupo social, la historia y tradiciones y otras que varían de comarca en comarca. Por lo tanto, no solamente es conveniente sino imprescindible investigar la validez y confiabilidad del instrumento antes de tomar decisiones a partir de los resultados obtenidos con su aplicación. El IESTA ha participado en este tipo de estudios.

En 2010 un proyecto orientado por investigadores del IESTA y de la Facultad de Psicología se dedicó a calibrar el test Wechsler Intelligence Scale for Children, edición 3 (WISC III) (Álvarez & Martínez, 2010).

“...compararemos al sujeto con la norma, o sea, con la actuación media o normal (en términos estadísticos). Arribar a este tipo de estudios es lo que se conoce como procedimiento de estandarización de la técnica. Así mismo es necesario que los resultados obtenidos puedan ser comparables, por lo que el registro del comportamiento observado debe ser llevado a un código lo más preciso y objetivo posible. Este aspecto refiere a la tipificación de los resultados. La tipificación implica uniformidad de procedimientos de aplicación y evaluación” (20).

El motivo de esta investigación fue que “a pesar del uso extendido del WISC III no se [poseían] datos normativos procedentes de nuestra propia población infantil” (49). Hasta ese momento los resultados del test se medían con “las tablas estadounidenses de conversión de puntajes incluidas en el manual que acompaña al material del test” (íd). El mérito adicional del proyecto fue que constituyó el “primer estudio de esta escala que se realiz[ó] con una muestra representativa de los escolares montevideanos” (50).

Con este antecedente, el IESTA estaba en condiciones de colaborar con la DGI. Es decir, en el caso de la DGI, los datos se manejaron a partir de esta experiencia. Sin embargo, no se trataba de validar el instrumento sino de manejar los resultados de acuerdo con el manual del contratista. No se trabajó en la evaluación de los resultados o en la contrastación con las tablas de los respectivos test.

El procesamiento de las respuestas, traducidas a valores cuantitativos en el lector óptico de la UAE, consistió en cargar estos datos en una base que se operó a partir de programar en un entorno digital tipo planilla de cálculo (*spreadsheet*) las relaciones múltiples de cada dato y con los factores propios del test. Es decir, se elaboraron los algoritmos que permitían llegar al índice de valoración de cada uno de los aspectos que analiza cada test. En esta actividad, el IESTA colaboró con personal informático de la DGI, quien realizó efectivamente el trabajo en su propio servidor.

A diferencia del proyecto con el WISC III, el IESTA no tomó conocimiento de los contenidos concretos de la prueba, sino de los valores asociados a ítems. Este aspecto fue especialmente estipulado en las coordinaciones iniciales. Mientras en 2010 todos los datos correspondían a sujetos anónimos, en este concurso había consecuencia personales para cada uno de los participantes y existían restricciones legales que afectaban al Tribunal. El IESTA podía satisfacer la necesidad de la DGI sin involucrarse con datos reservados o que tuvieran derechos propietarios.

El IESTA puede delimitar con mayor precisión su participación en proyectos que requieran sus recursos y competencias; de este modo puede alcanzar niveles de efectividad superiores.

Trabajo de campo

En este caso el trabajo de campo es toda la investigación preliminar para adecuar o alinear los recursos de la UAE y del IESTA y satisfacer las necesidades de la DGI. Hubo múltiples reuniones multipartitas. Por un lado, el Tribunal del Concurso con la UAE, más adelante se sumó el IESTA. Cuando se obtuvieron los resultados de la prueba de conocimiento, la DGI entendió que la FCEA estaba en condiciones de aportar una solución para recoger los datos de la prueba psicolaboral. Otra ronda de los tres actores mencionados, ahora con el agregado del contratista adjudicado para dicha prueba.

Las dos pruebas psicolaborales constituyeron diferentes desafíos para la UAE y para el IESTA. Ambas recogían datos que debían combinarse en diferentes fórmulas para alcanzar un resultado analizable por los psicólogos del contratista. Así, aparecía el reto de acceder a los algoritmos de la prueba sin revelar gratuitamente el *know how* original, pero permitiendo que el IESTA desarrollara los procesos necesarios para proveer información legible.

La figura 3 muestra un facsímil del formulario de la primera instancia.

DGI Universidad Católica del Perú
Llamado No. 0005/2015 DGI Administrativo Tributario

Formulario de respuesta
PRUEBA DE CONOCIMIENTO

POR FAVOR, ATIENDA LO SIGUIENTE
 Seguir las orientaciones de la prueba
 Rellenar cada círculo completamente ● ○ ○ ○ ○
 Para cambiar, borrar las marcas totalmente
 Usar lápiz No.2 o 2B preferentemente
 Rellenar el círculo correspondiente a la prueba y anotar el número en el cuadro a la derecha
 Rellenar el círculo correspondiente a su versión de la prueba y anotar el número en el cuadro a la derecha

Die	07	Año	2015
-----	----	-----	------

Prueba
 ① ② ③ ④ ⑤ ⑥

Versión
 ① ② ③ ④

RESPUESTAS

1	A	B	C	D	1
2	A	B	C	D	2
3	A	B	C	D	3
4	A	B	C	D	4
5	A	B	C	D	5
6	A	B	C	D	6
7	A	B	C	D	7
8	A	B	C	D	8
9	A	B	C	D	9
10	A	B	C	D	10
11	A	B	C	D	11
12	A	B	C	D	12
13	A	B	C	D	13
14	A	B	C	D	14
15	A	B	C	D	15
16	A	B	C	D	16
17	A	B	C	D	17
18	A	B	C	D	18
19	A	B	C	D	19
20	A	B	C	D	20
21	A	B	C	D	21
22	A	B	C	D	22
23	A	B	C	D	23
24	A	B	C	D	24
25	A	B	C	D	25

Figura 3. Facsímil del formulario para la prueba de selección

La propuesta de la prueba de conocimiento fue elaborada por el Tribunal del Concurso del sobre la base de 25 preguntas con cuatro opciones cada una. Había seis locaciones donde se realizaría la prueba y turnos, y por esa razón se generaron seis propuestas, cada una con cuatro versiones. Es decir, 24 pruebas diferentes. El apoyo de la UAE consistió en elaborar, siguiendo las orientaciones de las reuniones de coordinación, un formulario que cumpliera con las formalidades preestablecidas y facilitara la corrección. Es cierto que 5000 pruebas es un número que se ha manejado en pruebas académicas de la FCEA, pero en la etapa de planificación, la expectativa consideraba 12380 inscripciones.

Un elemento primordial, y que subyace a todas las pruebas tipo test con POM, es que la técnica para recoger los datos no constituya un obstáculo en sí misma. Se agregaron advertencias y se intentó que los aplicadores tuvieran la intervención mínima e imprescindible. Hay que considerar que un comentario extemporáneo puede sesgar los resultados entre diferentes ubicaciones. Se agregó el cuadro que aparece en la figura 4.

El trabajo de la FCEA, que se pactó al principio, terminaba con la extracción de las respuestas por aspirante, que se presentaron en una planilla de cálculo (*spreadsheet*, tipo Excel). El IESTA analizó la incidencia de las respuestas y se proporcionaron ambos informes.

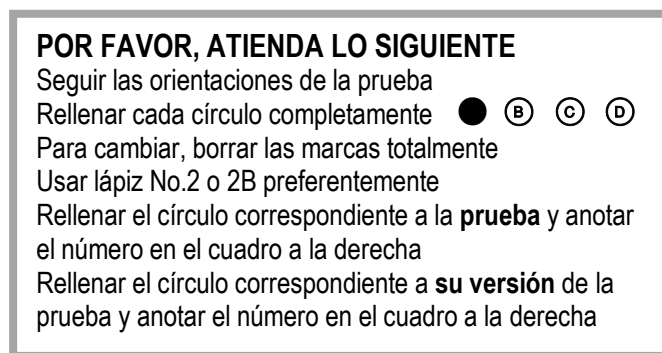


Figura 4. Instrucciones para responder la prueba de selección

La DGI solicitó apoyo para las dos pruebas psicolaborales y esa experiencia expandió la capacidad de la UAE para atender las necesidades internas en las UC de la FCEA. El producto, los nuevos formularios, demandaron ingenio y diseños inéditos, tanto por la necesidad de recoger datos con solvencia, como de aportar un reporte inicial - la planilla de cálculo -, que se adecuara al modelo de análisis diseñado por el IESTA. Es decir, un formulario que compatibilizara los requisitos de la prueba para responder según el diseño original del test; y que permitiera recuperar las respuestas en el formato de entrada de los datos al modelo del IESTA. Un desafío que sorteamos con paciencia y mucha coordinación.

El contratista de las pruebas psicolaborales utiliza instrumentos propios. Esto significa que las relaciones entre las preguntas o variables y la fórmula para obtener un resultado son propias. En ese sentido, elaborar un modelo de lectura que aporte la información para tomar decisiones sobre los perfiles de cada candidato; y, al mismo tiempo, sin revelar la ciencia detrás del modelo, demanda un tratamiento por lo menos peculiar. La solución a la que se arribó fue no tomar contacto con los contenidos de la propuesta, es decir, no acceder a las preguntas, pero sí al tipo de relación entre ellas: las diferentes combinaciones que deberían construirse con las respuestas de los 132 ítems. Esa actividad fue coordinada entre el contratista del componente psicolaboral y el IESTA. En este caso, el formato de informe mediante planilla de cálculo que produce normalmente la UAE resultó apropiado. La prueba de las piezas de dominó fue otra historia.

Los retos para administrar el test de dominó, y procesarlo en la UAE y el IESTA pueden resumirse en los siguientes: i) recoger las respuestas en un formato de formulario para POM, lo que facilitaba su lectura con el programa de la UAE; ii) que la anotación fuera accesible y no constituyera un obstáculo o una fuente de ansiedad para los concursantes; iii) que el formato de salida, el informe de la UAE, pudiera utilizarse directamente como insumo a la entrada del modelo de construcción de los resultados diseñado por el IESTA.

Primero se realizaron acuerdos y consultas con el contratista y con el IESTA. Los intercambios iniciales fueron en persona, luego hay documentadas 33 páginas de correspondencia en línea. El producto final – versión 9, en la historia de este proyecto – fue un formulario único, que incluía en la página uno la prueba psicolaboral con respuestas sobre una escala Likert y en la página dos el test de dominó, como se aprecia en la figura 5.

DGI DIRECCIÓN GENERAL DE INSTRUCCIÓN ADMINISTRATIVA
Llamado No. 0005/2015 DGI Administrativo Tributario

Formulario de respuesta PRUEBA PSICOTÉCNICA

Día: 11 Mes: Año: 2015

Número de concursante: [Grid]

ORIENTACIONES GENERALES
 Usar lápiz No.2 o 2B preferentemente.
 Respuesta correcta: (5) (4) ● (2) (1)
 Para cambiar, borrar las marcas totalmente.
 No doble ni arrugue este formulario.
 Respuesta incorrecta: (✓) (4) (2) (X)

Marque una sola respuesta en cada ítem

1-27 [Grids] 28-54 [Grids] 55-81 [Grids] 82-108 [Grids] 109-135 [Grids]

DGI DIRECCIÓN GENERAL DE INSTRUCCIÓN ADMINISTRATIVA
Llamado No. 0005/2015 DGI Administrativo Tributario

Señale el valor de la parte superior e inferior de la ficha de dominó faltante en cada gráfico

A Posición Respuesta correcta **B** Posición Respuesta correcta

1-14 [Grids] 15-32 [Grids] 33-50 [Grids]

UCE Unidad de Atención al Ciudadano FGA - UIR

Figura 5. Facsimil del formulario para recoger datos en las pruebas psicolaborales

En la primera página se repitieron las instrucciones de la prueba de selección, pero se refinó el texto a la vista de la experiencia con esa prueba. El cuadro se simplificó, ya habían sido expuestos a un formulario similar; y se agregaron orientaciones, como se muestra en la figura 6.

ORIENTACIONES GENERALES
 Usar lápiz No.2 o 2B preferentemente.
 Respuesta correcta: (5) (4) ● (2) (1)
 Para cambiar, borrar las marcas totalmente.
 No doble ni arrugue este formulario.
 Respuesta incorrecta: (✓) (4) (2) (X)

Figura 6. Orientaciones en el formulario de la prueba psicolaboral

El test de dominó requería orientaciones más elaboradas y desde una versión inicial donde se advertía en el encabezamiento “Señale el valor de la parte superior e inferior de la ficha de dominó faltante en cada gráfico”, llegamos a la versión 9 y definitiva. En la figura 7 aparece la versión 6. Las letras mayúsculas indican la pregunta en el formulario. Las de prueba tenían letras, estas cuatro, las del test se numeraron de 1 a 50.

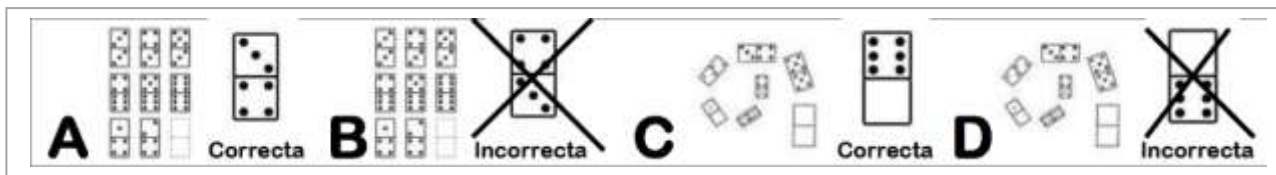


Figura 7. Orientación para el test de dominó, versión 6

El contratista entendió que los ejemplos para mostrar respuestas no incorrectas podrían inducir a error y se eliminaron. El espacio se utilizó para mostrar la notación en cada ítem y se reiteró en el espacio de respuestas reales, como se aprecia en la figura 8.

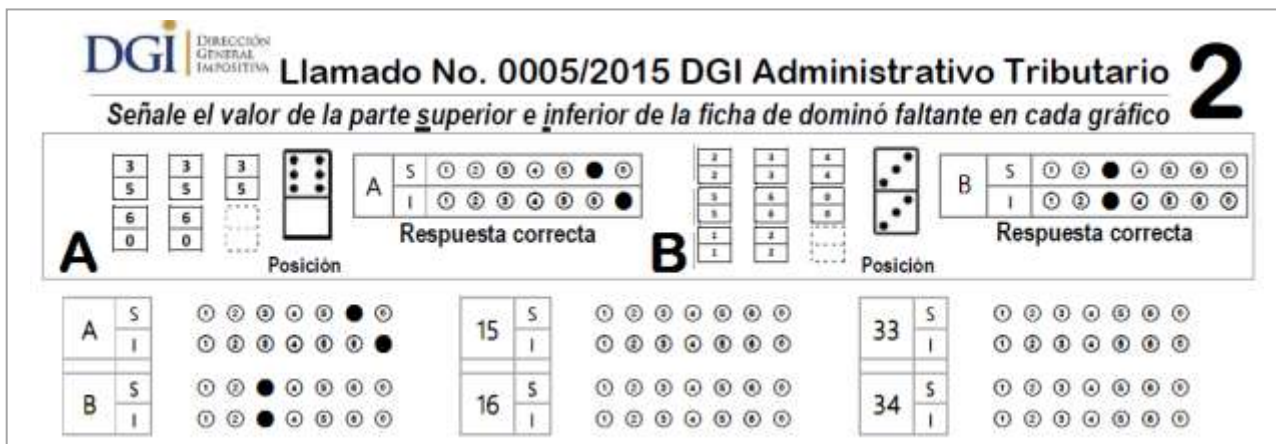


Figura 8. Instrucciones aplicadas al formulario del test de dominó

La prueba de selección fue aplicada en julio; las dos psicolaborales, el sábado 7 de noviembre. Se presentaron 904 candidatos. Para el día 11 ya estábamos estudiando los resultados y resolviendo las incongruencias. ¿Qué ocurre cuando no hay respuesta?

El modelo del IESTA requería respuesta en todos los ítems. Así, entre el contratista y el GTCI de la DGI acordaron, y la FCEA aplicó, el siguiente criterio:

“...acordamos que el escáner identifique como BLANK las preguntas que no fueron contestadas en la Prueba 1, de 132 enunciados en total. Luego a esas respuestas en Blanco, “BLANK”, se les asignaría un puntaje de 3, a los efectos de completar la hoja de respuestas de este test, ya que es la más neutral de los posibles enunciados “ni VERDADERO ni FALSO para mí” (Contratista, 12nov2015).

Más adelante fue necesario resolver algunos casos similares con el test de dominó. Una anécdota. Cuando se define que un dato es numérico, en oposición a literal, el programa no admite el cero a la izquierda, porque no tiene valor. De manera que si se intentaba señalar una ficha colocada con el valor 0 hacia arriba, o en primer término, por ejemplo, 02, esta notación contradecía la lógica del programa de lectura óptica. La resolvimos y todo resultó como estaba planeado. La solución fue discutida entre la UAE y el IESTA, con la DGI observando, y al final, acordada entre todos.

Emergentes del proyecto

Los emergentes del proyecto pueden describirse en dos áreas. Por una lado, lo que se esperaba obtener cuando se iniciaron las conversaciones con la DGI. En segundo lugar, lo que no se

esperaba, pero que apareció como consecuencia del involucramiento con la actividad. Innovaciones, solución de problemas, expansión de las competencias, por lo menos.

En el concurso

Los emergentes en el concurso se relacionan con el objeto del llamado propiamente dicho, es decir, la selección de los candidatos para cubrir las 68 plazas; y, por el otro, la evaluación del proceso, realizado con un procedimiento inédito para la DGI.

¿Cuál fue el balance de la realización de la prueba de conocimientos? Muy positivo. A pesar de que la afluencia fue inferior a las 12.000 personas habilitadas a dar la prueba (suceso que estadísticamente es esperable), la cantidad de personas que concurrieron permitió disponer de un universo suficiente de postulantes como para elegir los sujetos más adecuados para el puesto de trabajo. De un total de 4.854 concursantes habilitados el 77 % (**3.738** personas) alcanzó los 20 puntos por lo que pasaron a la siguiente instancia del concurso. Los 1.116 concursantes restantes no alcanzaron los 20 puntos en la prueba, siendo eliminados al no tener la posibilidad de alcanzar los 70 puntos, mínimo requerido para aprobar el concurso.

Asimismo la prueba se realizó dentro de los tiempos previstos y garantizó en todo momento la igualdad de condiciones entre los concursantes y preservó su anonimato.

¿Cuál fue el resultado de la aplicación del lector óptico en la corrección de las pruebas? La utilización del lector óptico para la corrección de las pruebas satisfizo ampliamente las expectativas del Tribunal del Concurso. Permitted arribar a los resultados en forma rápida y eficiente, garantizando la objetividad e imparcialidad y asegurando en todo momento el anonimato de los concursantes. Por otra parte permitió racionalizar los recursos humanos y materiales destinados a la corrección, lo que implicó una importante economía.

En la UAE

En 2014 se instaló el equipamiento para procesar formularios de test con POM en la UAE. En el imaginario original del proyecto se intentaría abarcar la mayor cantidad de UC con numerosidad, para facilitar el procesamiento de las evaluaciones de aprendizajes.

En esa línea, parecía conveniente tener un tipo de formulario estandarizado, tanto para facilitar el procesamiento, como para habituar a los estudiantes a un modelo que por consabido no interfería con el objetivo de las pruebas (Picón Martínez, 2014a). Es sabido que la forma de indagar los aprendizajes puede implicar dificultades adicionales que exceden las previstas en la propuesta e interfieren con el propio instrumento de evaluación. El ejemplo típico puede ser el ensayo.

Se pide a un estudiante que desarrolle determinado contenido, sea por memoria o por su aplicación a un caso ficticio – ad hoc para la prueba. El estudiante puede responder oralmente a un tribunal, como era de uso en otros tiempos, incluso con un bolillero que garantizaba el carácter aleatorio de las preguntas por donde se iniciaba el recorrido del programa completo. Con la masificación de la educación, y con la numerosidad en algunos centros, se dio paso a la instancia por escrito. Obstáculos adicionales en estos casos son la sintaxis y la ortografía. ¿Deben tener una ponderación en la calificación? Si así fuera, es posible que alguien terminara perdiendo el examen por su ignorancia de la gramática, que podría estar en las antípodas de los contenidos científicos de la asignatura en cuestión.

Con las pruebas que contiene POM puede pasar lo mismo. La máquina no delibera, el programa de lectura óptica contrasta la hoja de respuestas contra un programa preestablecido, que proyecta lo que debería encontrar en el formulario y cómo transformarlo en datos manejables para asignar una

calificación: nivel de aprendizaje constatado (Picón Martínez, 2014b). ¿Hay que penalizar errores en el manejo del formulario? Es el mismo caso de la gramática. Está aceptado que no, de manera que cuando el programa detecta una inconsistencia, no se contabiliza como error, sino que, pacientemente, se indaga la intención del estudiante. La mayor parte de las veces se constata que el estudiante no ha seguido las instrucciones del formulario: i) su elección no está señalada con carácter, la señal es tenue y no se detecta; ii) corrige una respuesta cambiando la configuración de la pregunta, borra marcas del formulario; iii) señala dos respuestas, pero indica la que vale con un lenguaje humano, como una tachadura. Esta experiencia sugirió que un único formulario para todas las UC venía a ser una restricción innecesaria.

A partir del trabajo para la DGI, la UAE ha podido diseñar formularios que se adecuan mejor a las propuestas de evaluación de cada UC. Puede decirse que proporcionó un mayor grado de audacia para la innovación. En realidad, nos abrió la cabeza muy mucho más allá de la propuesta del fabricante del programa. Por esa misma razón, el imaginario original, de expandir el uso de pruebas con POM dentro de la FCEA, tiene el potencial de un inventario abierto de opciones que pueden contemplar las expectativas de la mayoría, si no todas, las UC. Valga el siguiente ejemplo. Prueba de n ítems con cuatro opciones, donde hay de una a cuatro respuestas y uno, dos o tres distractores. Es decir, que en algunos ítems el estudiante debería señalar una opción; en otros, dos; y hasta en algunos, tres y cuatro. La UAE ha diseñado un formulario adaptado a ese modelo de pregunta y que puede procesarse con el lector óptico, para proporcionar los datos que necesita la UC.

Además, la UAE ha quedado posicionada como fuente para el suministro de este servicio fuera de la FCEA. No es el único lector óptico de la UDELAR, por supuesto, pero tiene detrás un equipo que se adapta a las necesidades, puede abordar los problemas con solvencia y propone soluciones factibles. Todo lo que se espera de una unidad de apoyo.

Comentario final y recomendaciones

Desde el comienzo, dada la cantidad de postulantes que se presentaron al Llamado 0005/2015 y ante la necesidad imperiosa de llenar los puestos vacantes el Tribunal del Concurso fue consciente de que para culminar el proceso de selección tendría que diseñar pruebas y utilizar procedimientos de aplicación y corrección rápidos, simples y económicos; se enfrentaba a limitaciones de tiempo y presupuestales.

Surge allí la idea de acudir a la FCEA. Para la prueba de conocimiento se vislumbró desde el inicio la utilidad de utilizar POM; las dudas se presentaron principalmente para la elaboración, toma y corrección de la evaluación psicolaboral.

La numerosidad hizo imposible la aplicación de entrevistas a los candidatos: no se contaba con el tiempo. A través de las entrevistas se establece una comunicación interpersonal que permite obtener información para corroborar las hipótesis generadas durante el proceso de evaluación laboral.

Es por ello que se decidió incorporar a la batería a utilizar el test proyectivo de Wartegg con la finalidad de comprobar la evaluación diagnóstica hasta el momento realizada y eventualmente detectar elementos psicopatológicos.

De esa manera se pudo completar la evaluación psicolaboral con técnicas diagnósticas a las cuales las POM les imprimió tanto en la ejecución y como en la corrección, velocidad, economía y exactitud.

En suma, la utilización de POM y procesamiento de lectura óptica permitió cumplir eficientemente con los plazos, con resultados rápidos, económicos y por sobre todo confiables.

El proyecto de apoyo a la DGI con motivo de la convocatoria 5-2015 puso en evidencia el potencial de la FCEA para satisfacer necesidades de otros actores institucionales de la Administración, mediante recursos que inicialmente no fueron establecidos para la extensión y las relaciones con el medio. Además, fue una oportunidad para coordinar diferentes unidades académicas (UA), como el IESTA y la UAE, que produjo una sinergia inédita. Cada experiencia de relaciones con el medio es una oportunidad de aprendizaje y de servicio a la comunidad. Otras UA podrían encontrarse en una situación similar, por lo cual sería recomendable que a su tiempo realizaran una prospección sobre los potenciales escenarios donde su contribución, desde la UDELAR, podría fortalecer sus competencias para la investigación, la creación de conocimiento y la enseñanza.

Los aprendizajes durante el proyecto se manifiestan en la calidad del apoyo a las UC, en el caso de la UAE, con alto grado de efectividad; de manera que los responsables de las UC trasladan sus expectativas para diseñar instrumentos de recolección de POM, en lugar de aceptar pacíficamente la imposición de los formularios estandarizados que ofrece el fabricante del programa de lectura óptica. Diseño e innovación.

Bibliografía y referencias

- Alles, M. (2010). *Selección de Competencias*. Buenos Aires: Granica.
- Álvarez, R., & Martínez, s. (2010). *El WISC III en los escolares. Baremo Montevideo*. Montevideo: Waslala.
- Bloom, B. S. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals*. Philadelphia: David McKay Company.
- Davidson, C. (2011a). *Now You See It: How the Brain Science of Attention Will Transform the Way We Live, Work, and Learn*. New York: Viking.
- Davidson, C. (23 de 09 de 2011b). *Standardized tests for everyone? In the Internet age, that's the wrong answer*. Recuperado el 6 de setiembre de 2016, de www.washingtonpost.com: https://www.washingtonpost.com/opinions/standardized-tests-for-everyone-in-the-internet-age-thats-the-wrong-answer/2011/09/21/gIQA7SZwqK_story.html?utm_term=.21035ad6888e
- DGI. (2015). *Llamado No.0005/2015 Administrativo Tributario - MEF - Dirección General Impositiva*. Obtenido de Uruguay concursa: <https://www.uruguayconcurso.gub.uy/Portal/servlet/com.si.retsel.verllamado?5625>
- Kamenetz, A. (2015). *The Test: Why our schools are obsessed with standardized testing - but you don't have to be*. Philadelphia: PublicAffairs.
- Kelly, F. J. (feb de 1916). The Kansas Silent Reading Tests. *Journal of Educational Psychology*, 8(2), s/d.
- ONSC. (2015). *Manual de Reclutamiento y Selección 2015*. Montevideo: ONSC.
- Picón Martínez, A. (2014a). *Diseño de formularios de test para lector óptico. Documento interno*. Montevideo: FCEA-UAE.
- Picón Martínez, A. (2014b). *Diseño de matrices para lector óptico. Documento interno*. Montevideo: FCEA-UAE.
- Picón Martínez, A. (2016). Evaluación con preguntas de opción múltiple. Desafíos y oportunidades (en prensa). (pág. <http://www.ccee.edu.uy/jacad/2016/index.html>). Montevideo: FCEA.
- Rodríguez Ayán, M. N. (2014). *Diseño y análisis de pruebas de múltiple opción*. Montevideo: UDELAR-FQ.

- Spencer, L., & Spencer, S. (1993). *Competence at Work. Models for Superior Performance*. Hoboken, NJ: John Wiley & Sons.
- Thorndike, E. (1903). *Educational Psychology*. New York: The Science Press.
- Thorndike, E. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, 4(1), 25-29.
- UDELAR. (2011). *Ordenanza de estudios de grado y otros programas de formación terciaria*. Ordenanza, Universidad de la República, Consejo Directivo Central, Montevideo.
- UDELAR. (2012). *Facultad de Ciencias Económicas y de Administración. Reglamento del Plan de estudios*. Reglamento, Universidad de la República, Consejo Directivo Central, Montevideo.
- UDELAR-FCEA. (26 de noviembre de 2013). Reglamento de Concursos para la Provisión de Cargos Docentes de la Facultad de Ciencias Económicas. *Resolución CDC No.29*. s/d.
- Watters, A. (27 de 01 de 2015). *Multiple Choice and Testing Machines: A History*. Recuperado el 6 de set de 2016, de <http://hackededucation.com>: <http://hackededucation.com/2015/01/27/multiple-choice-testing-machines>
- Weigle, C. (1996). *El Psicólogo evaluador en el campo laboral*. Buenos Aires: Autor.