DISCUSIONES ACTUALES SOBRE EL USO DE INSTRUMENTOS DE MEDICIÓN PSICOLÓGICA EN PROCESOS DE EVALUACIÓN

En la actualidad, la evaluación psicológica y el desarrollo de instrumentos de medición han experimentado una diversificación, no solo en cuanto a la **población objetivo** (paso de evaluaciones individuales a evaluaciones masivas de estudiantes en los diferentes niveles educativos, así como de personal que aspira a ingresar a cargos de carrera administrativa en el sector público y privado), sino también en cuanto a los **atributos** que se pretenden medir (paso de énfasis cognitivo, centrado en los conocimientos y competencias en áreas fundamentales de la educación básica y media, al análisis de competencias funcionales y comportamentales que se espera despliegue los docentes y directivos docentes en su quehacer pedagógico, por citar solo un ejemplo), y a la **finalidad** de dichas evaluaciones (diagnóstico, selección, clasificación, ascenso, reubicación, nombramiento en propiedad) (Anitei & Buzea, 2012).

Colombia no ha sido ajena a esta dinámica, de manera que se pueden apreciar tres tendencias importantes en lo que a procesos de evaluación masiva se refiere, procesos que, en sus diferentes fases, hacen uso de instrumentos psicológicos:

1. La reglamentación e implementación del Sistema Nacional de Evaluación de la Calidad de la Educación, en donde, a través del ICFES, se aplican anualmente evaluaciones censales a los estudiantes de 3, 5 y 9 grado (pruebas SABER 3, 5 y 9), grado 11 (SABER 11) y estudiantes de últimos semestres de educación superior (SABER PRO), donde se pretende analizar y hacer seguimiento al desarrollo de las competencias

Ana Cristina **Santana Espitia** / acsantanae@unal.edu.co
Estudiante del Doctorado en Psicología
Universidad Nacional de Colombia, sede Bogotá

básicas de los estudiantes en sus diferentes niveles educativos, en concordancia con los estándares básicos de competencias de las áreas que son evaluadas por las pruebas (Grupo de Investigación sobre Pruebas Masivas, 2008; Hoyos, 2010).

2. La promulgación de la Ley 909 de 2004, la cual consagra la estructura, organización y funciones de la Comisión Nacional del Servicio Civil (CNSC), entidad que, de acuerdo con el Artículo 130 de la Constitución Política de Colombia, es responsable de la administración y vigilancia de las carreras de los servidores públicos. Está orientada a

> Posicionar el mérito y la igualdad en el ingreso y desarrollo del empleo público; velar por la correcta aplicación de los instrumentos normativos y técnicos que posibiliten el adecuado funcionamiento del sistema de carrera; y generar información oportuna y actualizada, para una gestión eficiente del sistema de carrera administrativa. (Comisión Nacional del Servicio Civil, 2014)

La Ley 909 contempla también lo concerniente a empleo público, gerencia pública y normas de carrera administrativa

3. Específicamente, en el sistema de carrera para los docentes que se encuentran vinculados al sector público, mediante el Decreto 1278 de 2002, se ha incorporado la cultura de la evaluación en diferentes momentos del servicio docente: ***evaluación de ingreso,*** donde el aspirante a docente es evaluado en aptitud numérica, aptitud matemática, saber pedagógico, dominio sobre los conocimientos profesionales disciplinares básicos y prueba psicotécnica (actitudes, intereses y motivaciones que se ponen en juego cuando se abordan situaciones propias de procesos pedagógicos o de gestión institucional) (Instituto Colombiano para la Evaluación de la Educación, 2013); ***evaluación de período de prueba y de desempeño***, en donde se evalúan competencias funcionales y comportamentales relacionadas con el cargo, y ***evaluación de competencias***, dirigida a los docentes y directivos docentes que deseen ascender o reubicarse dentro del escalafón.

Las discusiones contemporáneas sobre el uso de los instrumentos psicológicos, específicamente, en evaluaciones masivas en educación, aunque bien podrían ser aplicables en otras esferas de la sociedad donde la evaluación tiene lugar, se han centrado, principalmente, en la adecuación y pertinencia de las pruebas para evaluar a individuos con diferencias en cuanto a estrato socioeconómico, cultura, etnia, región de origen, ya que se aduce que las pruebas favorecen a los individuos de grupos específicos (ej. personas de estratos altos), mientras que las personas no pertenecientes a dichos grupos obtienen resultados inferiores a los esperados. Al respecto, Bustamante (2003) plantea, en su artículo, una serie de hipótesis sobre el uso de las evaluaciones masivas en Colombia, desde la perspectiva del sociólogo francés Pierre Bourdieu. Para efectos de este ensayo, se analizarán cuatro de ellas (las hipótesis 5, 6, 8 y 13), a fin de determinar sus pros y contras en el marco de la evaluación masiva nacional y, específicamente, en cuanto al uso de instrumentos de medición psicológica.

La quinta hipótesis de Bustamante (2003), señala que "***Si estas evaluaciones masivas entienden que la apropiación simbólica es igual para todos, que todos se benefician igualmente de ella, posiblemente poco ayuden a esclarecer los efectos estructurales de la desigualdad ligados al uso del lenguaje***" (p. 36;énfasis agregado). En este orden de ideas, el autor menciona que no es posible garantizar el supuesto de homogeneidad en las condiciones de aplicación de las pruebas o la tenencia del atributo objeto de evaluación de la misma manera para todos los individuos, dado que el lenguaje que se usa en la construcción de las pruebas reproduce las desigualdades que se observan en la estructura social, y "como la apropiación simbólica se da a través de la estructura social, y como ésta no es igualitaria, por lo tanto, no todos tienen acceso a lo simbólico de la misma manera, ¡así se les

haya dicho lo mismo!" (p. 36), de manera que, en términos del autor, "la evaluación, más allá de medir algo, hace aparecer como diferencia de competencia individual (antes se decía "inteligencia") lo que ha sido causado socialmente" (p. 36).

Si bien es cierto que los usuarios de las evaluaciones masivas son individuos con diferencias culturales y geográficas importantes, no quiere decir que posean esquemas conceptuales diferentes por su pertenencia a dichos grupos, es decir, la manera como los individuos se apropian y organizan la información de las diferentes áreas del conocimiento o en la definición de los atributos y características que configuran su personalidad tienen aspectos comunes, con independencia de su pertenencia a uno u otro grupo social, comunalidades que hacen posible la comunicación entre los individuos y la difusión del conocimiento.

En ese sentido, es preciso reconocer, no obstante, las comunalidades en los esquemas conceptuales, que ***existen acercamientos diferenciales a los objetos de la cultura por parte de los individuos***, y estas diferencias no obedecen a criterios psicométricos o de construcción de los instrumentos, de modo que Bustamante (2003) sitúa como un problema de las evaluaciones aquello que es un problema de la manera como las personas acceden al conocimiento, bien sea a los instrumentos de producción de las competencias y a los lugares de expresión legítima (contextos de aplicación).

Teniendo en cuenta lo anterior, lo que el sistema educativo debe garantizar es que todos los individuos tengan la misma disponibilidad y acceso a los objetos de la cultura en general, no solo los que son susceptibles de ser evaluados en estrategias de lápiz y papel. El contenido de los instrumentos de medición debe responder, precisamente, a esos objetos culturales y educativos que son aprehendidos por la totalidad de los estudiantes colombianos en el aula e, incluso, en las interacciones que se dan en el aula y fuera de ella (p. ej. piénsese el caso de las competencias ciudadanas). Al respecto Álvarez (2001), indica que:

> Entendida la educación como acceso a la cultura y a la ciencia –bienes comunes históricamente construidos–, el reto que cada profesor tiene es no dejar a nadie fuera. Tomar conciencia de este hecho es comprometerse con modos razonables de actuar con cada sujeto que se encuentre en esa situación delicada para no excluir a nadie de la participación del saber. (p. 13)

La sexta hipótesis de Bustamante (2003) menciona que "*si estas evaluaciones masivas coadyuvan a construir la idea de una lengua 'común', su acción puede haber ido en detrimento de expresiones lingüísticas y comunicativas que son significativas desde el punto de vista social*" (p. 37; énfasis agregado). En este apartado, el autor menciona que la escuela y las evaluaciones masivas privilegian una lengua oficial particular, una manera lingüística concreta en la cual están estructurados tanto los contenidos educativos como la arquitectura formal de las evaluaciones aplicadas en el aula y fuera de ella, generando con ello, en términos del autor, un desconocimiento y rechazo de las expresiones lingüísticas como dialectos y regionalismos provenientes del saber cultural de grupos considerados como minoritarios.

Con relación a esta hipótesis, puede mencionarse que el autor presenta una confusión entre los criterios internos del uso de un conocimiento público, como lo son los criterios internos de una lengua, con las variaciones culturales válidas, reconocidas para comunidades minoritarias. Toda área del conocimiento tiene una ***estructura pública***, que debe ser conocida por los usuarios de la misma, y la evaluación no puede ser ajena a las características formales y de contenido de dichas áreas, de manera que ***los instrumentos de medición son también objetos culturales públicos*** que no pueden violar los criterios públicos a los cuales deben obedecer; eso no significa que se ignoren las variaciones culturales pertinentes.

La octava hipótesis de Bustamante (2003) indica que "*si estas evaluaciones masivas fetichizan y relativizan el lenguaje, posiblemente contribuyan poco a entender el valor de uso de la lengua*" (p. 37; énfasis agregado). En esta hipótesis, el autor menciona que, aunque en las evaluaciones educativas bien sea dirigidas a estudiantes o docentes se definan las competencias como un "saber hacer en contexto", entendiendo el contexto como el brindado por los textos que se emplean para elaborar las preguntas, los instrumentos de evaluación son el contexto donde solo se evalúan las áreas de conocimiento en las situaciones creadas por la prueba.

En cuanto a esta observación de Bustamante (2003), es importante señalar que la elaboración de instrumentos de medición psicológica de cualquier atributo sigue por un proceso riguroso de construcción en donde participan docentes expertos en las áreas de conocimiento (docentes, para el caso de las evaluaciones masivas de corte cognitivo), o expertos en las competencias funcionales y comportamentales (docentes y directivos docentes con vasta experiencia en el sector educativo, para el caso de las pruebas docentes en sus diferentes etapas), de modo que el contenido de los ítems que conforman las

pruebas no es artificial y arbitrario, sino producto de consensos entre profesionales expertos, donde hay acuerdos no solo en la naturaleza de las temáticas a evaluar, sino que las pruebas se sustentan en procesos de validación de constructo y de contenido que garantizan la calidad de las inferencias que pueden realizarse a partir de los resultados obtenidos por las pruebas.

En ese sentido, Gómez e Hidalgo (2005) citan la posición de Messick respecto a la validez señalando que

> [...] la validez no es del test o de la observación, lo que se valida son las inferencias derivadas de las puntuaciones del test o de otros indicadores, inferencias sobre el significado de las puntuaciones o la interpretación para propósitos aplicados y sobre las implicaciones para la acción, es decir, las consecuencias sociales y éticas. (p. 5)

La hipótesis 13 de Bustamante (2003), menciona que "***por su propio funcionamiento, estas evaluaciones masivas pueden haber contribuido a unificar criterios curriculares, pedagógicos y lingüísticos***" (p. 40; énfasis agregado). El autor menciona que a partir de los resultados de las evaluaciones masivas, pueden extraerse cuatro tipos de validez:

1. Validez de los usos lingüísticos, en términos de las respuestas correctas e incorrectas en pruebas nacionales de tipo censal o muestral.
2. Validez de las prácticas pedagógicas, en cuanto al fomento y promoción de aquellas que contribuyen a lograr buenos desempeños en las pruebas.
3. Validez de las costumbres educativas, ya que la cultura de la evaluación orienta la planeación y ejecución de los elementos curriculares (planes de área, de asignatura) en todas las áreas y grados.
4. Validez de las políticas educativas, pues tienen valor aquellas que apuntalen y sean direccionadas por las evaluaciones masivas.

Con relación a esta hipótesis, no puede desconocerse que las inferencias resultantes de los procesos evaluativos han sido determinantes en la toma de decisiones en distintos niveles del sistema educativo y, en ese sentido, los instrumentos de medición (pruebas) son solo un insumo, no la única herramienta que se toma en cuenta para el diseño de políticas educativas a nivel local y nacional. Por ello, es tan importante garantizar la validez de los instrumentos de medida, porque lo que entra en juego es "la adecuación, significado y utilidad de

las inferencias específicas hechas con las puntuaciones de los tests" (APA, AERA, NCME, 1985, p. 8). En la determinación de la validez de un instrumento de medición,

> Una variedad de evidencias pueden obtenerse de las puntuaciones producidas por un test dado, y hay muchas formas de acumular evidencia para apoyar una inferencia específica. (...) Aunque la evidencia puede ser acumulada de muchas formas, la validez se refiere siempre al grado en que esa evidencia apoya las inferencias que se hacen a partir de las puntuaciones. (APA, AERA, NCME, 1985, p. 8).

Específicamente, la hipótesis de Bustamante (2003) señala 4 tipos de evidencia, los cuales pueden extraerse a partir de los resultados de las evaluaciones masivas a nivel educativo, las cuales deben ser analizadas cuidadosamente, para determinar su relevancia y pertinencia dentro del sistema educativo. En cuanto a la validez de los usos lingüísticos, ya se comentó, en un apartado anterior, que no se ignoran las variaciones culturales y lingüísticas de los grupos denominados minoritarios, sino que lo que pretende evaluar la prueba, en tanto objeto cultural público y accesible para todos los usuarios (estudiantes, en este caso), es un conjunto público de conocimientos y/o dimensiones cognitivas que configuran las áreas del saber, las cuales deben ser asequibles para todos los estudiantes. Por ello, no se evalúa como correcto o incorrecto el saber cultural de los estudiantes, sino se valora su desempeño en un producto público, que recoge una cierta evidencia de los aprendizajes en el aula, complementando las evidencias provenientes del proceso de evaluación formativa que se da en el aula de clase (Álvarez, 2001).

La validez de las prácticas pedagógicas, a diferencia de lo planteado por Bustamante (2003), no se garantiza exclusivamente por los resultados de las pruebas externas. Las prácticas pedagógicas tienen su mayor evidencia de validez en la medida en que los estudiantes aprehendan los conocimientos y competencias propias del área; además, que estén en la capacidad de utilizar ese conocimiento en situaciones de la vida cotidiana, aspecto que puede comprobarse principalmente en los procesos de evaluación interna e, incluso, en las interacciones escolares y aquellas que trasciendan el espacio físico de la institución.

Los resultados de las pruebas educativas no deben usarse para hacer procesos de entrenamiento mecánico en las instituciones educativas ni deben ser empleadas como mecanismo sancionatorio de X o Y abordaje pedagógico, el ***quid*** del asunto es lograr que los docentes se acerquen a los resultados de las pruebas externas no con temor y

recelo, sino que se haga un ***uso pedagógico de los resultados de dichas evaluaciones***, en la medida en que permitan retroalimentar no solo las competencias y componentes en los que los estudiantes presentan altos y bajos desempeños, sino proporcionen pistas acerca de las posible continuidad o adecuación de aspectos pedagógicos, metodológicos y didácticos de la clase que podrían redundar en el desarrollo académico y convivencial de los estudiantes.

Los docentes no deben enseñar e impartir conocimientos en función de los resultados de las pruebas externas y/o de las posibles consecuencias asociadas a ellas (evaluación de altas consecuencias) (Tierney, 2014), sino enseñar para formar mejores individuos en todos los aspectos, siendo la escuela el escenario privilegiado, donde tienen lugar los procesos de socialización que iniciaron desde el seno familiar. Álvarez (2001) define, de manera precisa, lo que debe ser la evaluación en educación: "En el ámbito educativo debe entenderse la evaluación como actividad crítica de aprendizaje, porque se asume que la evaluación es aprendizaje en el sentido que por ella adquirimos conocimiento" (p. 12).

En cuanto a la validez de las costumbres educativas, es preciso indicar que la cultura de la evaluación no viene dada única y exclusivamente por los procesos de evaluaciones masivas. Cualquier proyecto educativo institucional y, en consonancia, cualquier modelo pedagógico que adopta una institución educativa, tiene de fondo una concepción de evaluación, de lo que se considera debe ser el objeto de la misma, inclusive, de las estrategias más eficaces para acceder a las evidencias que permitan dar cuenta del grado de aprehensión de dichos objetos evaluativos. La evaluación también, transversaliza, por así decirlo, los procesos mismos de gestión institucional, tal es el caso de los Planes de Asignatura y Área, Planes de Mejoramiento Institucional, Planes de Apoyo al Mejoramiento, Planes Operativos Anuales.

El análisis de los desempeños docentes, a su vez, es visto desde la lupa de la evaluación de desempeño para los docentes regidos por el Decreto 1278 de 2002, en el que se analizan las competencias comportamentales y funcionales de cada cargo, frente a las evidencias que aporta el docente (rúbricas, portafolios, proyectos pedagógicos, observaciones de clase). Para este caso, aun cuando las competencias funcionales y comportamentales vienen dadas a partir de las directrices del Ministerio de Educación Nacional y la Comisión Nacional del Servicio Civil, hay consensos, desde principio de año, entre los docentes y los evaluadores sobre las actividades a desarrollar y los productos que serán tenidos en cuenta como objetos de evaluación.

Finalmente, con relación a la validez de las políticas educativas, cabe anotar que las inferencias provenientes de los resultados de las evaluaciones masivas en educación aplicadas tanto a estudian-

tes como a docentes han tenido una considerable incidencia no solo en cuanto a la adopción de políticas públicas (ej. Promulgación de la Ley 1620 de 2013 de convivencia escolar, inclusión del componente de competencias ciudadanas en las pruebas SABER), sino en la estructuración de programas e iniciativas educativas diseñadas con el propósito de fortalecer las competencias de los estudiantes (Programa Todos a Aprender, Programa de Bilingüismo, promoción de mecanismos de convivencia ciudadana y defensa de los derechos humanos, sexuales y reproductivos). A nivel docente, los resultados de las evaluaciones masivas constituyen insumos para el desarrollo de Programas de Formación Docente, en aras de una cualificación permanente en su quehacer pedagógico.

Una vez revisadas estas hipótesis, se aprecia la necesidad de continuar indagando sobre la naturaleza de las implicaciones sociales, éticas y económicas en la aplicación de instrumentos psicológicos en procesos de evaluación a gran escala. Si bien se han dado discusiones importantes en el sector educativo, en donde la tendencia general apunta al reconocimiento de la importancia de los procesos evaluativos tanto de procesos como de los miembros del sistema educativo, pese a las voces de oposición de ciertos gremios docentes y sindicales (Hoyos, 2010), es importante analizar la incidencia del uso de las pruebas en escenarios como la selección de personal administrativo y de carrera en otro tipo de organizaciones y esferas estatales (ej. rama judicial, organismos adscritos a ministerios como el de Salud, de Protección Social).

El debate también puede darse en torno al uso de atributos psicológicos como personalidad, intereses, actitudes (Rothstein & Goffin, 2006), e incluso trascender las técnicas de evaluación tradicionales, como las pruebas de lápiz y papel, de modo que otras técnicas de aproximación al conocimiento psicológico de los individuos, como las entrevistas situacionales y la estructura que deben tener en esta clase de procesos de selección, sean aspectos tenidos en cuenta por psicólogos y psicómetras (Whetzel & McDaniel. 2009).

Referencias

Álvarez, J. (2001). El campo semántico de la evaluación: Más allá de las definiciones. En: J. Álvarez (Ed.). ***Evaluar para conocer, examinar para excluir*** (pp. 11-26). Madrid: Ediciones Morata

American Psychological Association [APA], American Educational Research Association [AERA], and National Council on Measurement in Education [NCME]. (1985). ***Standards for educational and psychological testing***. Washington, D.C.: American Psychological Association

Anitei, M. & Buzea, E. (2012). Models and applications in psychological selection and evaluation. ***Procedia - Social and Behavioral Sciences***, ***33***, 756-760.

Bustamante, G. (2003). Competencias y evaluaciones masivas en Colombia: Una mirada desde Bourdieu. ***Pedagogía y Saberes***, (18), 33-44.

Comisión Nacional del Servicio Civil (2014). ¿Quiénes somos? Recuperado de http://www.cnsc.gov.co/index.php/institucional/direccionamiento-estrategico/quienes-somos-cnsc

Gómez, J. & Hidalgo, M. (2005). La validez en los tests, escalas y cuestionarios. ***Revista La sociología en sus escenarios, 12,*** 1- 14.

Grupo de Investigación sobre Pruebas Masivas en Colombia. (2008). El surgimiento del Servicio Nacional de Pruebas del ICFES en las voces de sus protagonistas. ***Magistro, 2***(3), 115-134.

Hoyos, L. (2010). Evaluaciones masivas y estandarizadas, mal necesario, para medir la calidad de la educación en Colombia. ***Pedagogía Magna, 8,*** 108-119.

Instituto Colombiano para la Evaluación de la Educación [ICFES]. (2013). ***Guía de orientación para las pruebas de aptitudes y competencias básicas y psicotécnica de ingreso a la carrera docente para directivos docentes y docentes de población mayoritaria***. Bogotá: ICFES

Rothstein, M. G. & Goffin, R. D. (2006). The use of personality measures in personnel selection: What does current research support? ***Human Resource Management Review***, ***16***(2), 155-180.

Tierney, R. D. (2014). Fairness as a multifaceted quality in classroom assessment. ***Studies in Educational Evaluation.*** Recuperado de: http://dx.doi.org/10.1016/j.stueduc.2013.12.003

Whetzel, D. L. & McDaniel, M. A. (2009). Situational judgment tests: An overview of current research. ***Human Resource Management Review,*** 19(3), 188-202.