



Revista de Ciencias Sociales (Ve)
ISSN: 1315-9518
rcs_luz@yahoo.com
Universidad del Zulia
Venezuela

Prueba adaptativa para medir el progreso de los estudiantes: Colegios en Bogotá, Colombia

Orjuela Roa, Cristian Harrison; Lesmes Martínez, Óscar Mauricio; Tarazona G., Nicolas Marcel; Castañeda Polanco, Juan Gabriel

Prueba adaptativa para medir el progreso de los estudiantes: Colegios en Bogotá, Colombia

Revista de Ciencias Sociales (Ve), vol. 26, 2020

Universidad del Zulia, Venezuela

Disponible en: <https://www.redalyc.org/articulo.oa?id=28064146023>



Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial 3.0 Internacional.

Prueba adaptativa para medir el progreso de los estudiantes: Colegios en Bogotá, Colombia

Adaptive test to measure student progress: Schools in Bogotá, Colombia

Cristian Harrison Orjuela Roa
Ceinfes, Colombia
herramientas@ceinfes.com

Redalyc: <https://www.redalyc.org/articulo.oa?id=28064146023>

Óscar Mauricio Lesmes Martínez
Ceinfes, Colombia
oscar.lesmes@beplusgrouppla.com

Nicolas Marcel Tarazona G.
Ceinfes, Colombia
ceinfes.pyd.lenguaje@gmail.com

Juan Gabriel Castañeda Polanco
Corporación Universitaria Minuto de Dios, Colombia
jucastaneda@uniminuto.edu

Recepción: 12 Marzo 2020
Aprobación: 28 Mayo 2020

RESUMEN:

En el ámbito escolar la pérdida de fiabilidad de la evaluación ha venido constituyéndose en una problemática cada vez más evidente. En ese sentido, esta investigación tuvo como objetivo analizar los resultados de aplicación de una prueba adaptativa, en estudiantes de ciclo III de algunos colegios de Bogotá, Colombia. Para lo cual, se utilizó un diseño metodológico cuantitativo-transeccional con alcance correlacional, y se seleccionó una muestra intencionada no-probabilística. El instrumento de prueba correspondió al de una técnica Rasch, con modelo logístico de un parámetro asociado a la teoría de respuesta al ítem, desde el cual se obtuvieron los datos de desempeño de 544 estudiantes, los cuales fueron interpretados con base en el análisis estadístico. Los resultados evidenciaron que: El test diseñado pudo identificar y adaptarse a las características de los estudiantes considerando los criterios curriculares de evaluación; entre los hallazgos de los estudiantes no se muestran diferencias significativas que indiquen algún factor de inestabilidad en la prueba aplicada. Se concluye, que la aplicación de esta prueba en Instituciones Educativas de Colombia, puede ayudar a fortalecer componentes de la evaluación escolar y progreso de los alumnos, frente a los aprendizajes esperados para el proceso, si se enfocan en desempeños de competencia.

PALABRAS CLAVE: Prueba adaptativa, evaluación, prueba computarizada, estudiantes, Colombia.

ABSTRACT:

In the school environment, the loss of reliability of the evaluation has become an increasingly evident problem. In this sense, this research aimed to analyze the results of the application of an adaptive test, in students of cycle III of some schools in Bogotá, Colombia. For which, a quantitative-transectional methodological design with correlational scope was used, and an intentional non-probabilistic sample was selected. The test instrument corresponded to that of a Rasch technique, with a logistic model of a parameter associated with the item response theory, from which the performance data of 544 students were obtained, which were interpreted based on the statistical analysis. The results showed that: The designed test could identify and adapt to the characteristics of the students considering the curricular evaluation criteria; There are no significant differences among the students' findings that indicate some factor of instability in the applied test. It is concluded that the application of this test in Educational Institutions of Colombia can help to strengthen components of the school evaluation and student progress, compared to the expected learning for the process, if they focus on competency performances.

KEYWORDS: Adaptive test, evaluation, computerized test, students, Colombia.

INTRODUCCIÓN

En el ámbito escolar la pérdida de fiabilidad de la evaluación ha venido constituyéndose en una problemática cada vez más evidente. Y esto no solo por la falta de intencionalidad en algunas de sus prácticas, sino sobre todo por la ausencia de una retroalimentación educativa, clara y precisa, sobre las debilidades y fortalezas de los estudiantes desde los procesos de evaluación de aprendizajes.

En este sentido, las reflexiones sobre la problemática evaluativa escolar han venido mostrando que existe un uso artificioso del proceso de evaluación, que deriva, por una parte, en invisibilizar dentro de la comunidad escolar los déficits en la calidad de los aprendizajes (Rodríguez, Muñoz y Castillo, 2014; Cortes, Daza y Castañeda, 2019); por otra, en generar una serie de cuestionamientos pedagógicos en torno a la pertinencia de la evaluación en los procesos educativos (Celman, 2005; Shepard, 2006; Rodríguez, et al., 2014). Lo que, en general, comprueba la gravedad de la situación.

Aunque medir el progreso alcanzado por los estudiantes en sus aprendizajes no constituye por sí misma, una expresión de la calidad de la educación, sí existe hoy, al menos un consenso sobre la necesidad y utilidad de la evaluación para dar cuenta de las demandas de cambio en los procesos de enseñanza (Yovane, 2000). Es decir, que –a pesar de no ser un factor directo de la calidad educativa- la evaluación deviene fundamental en la actualidad, en cuanto es la que permite evidenciar el nivel de desempeño estudiantil, con base en el cual se identifican los aspectos que necesitan atención para que se dé el desarrollo de competencias (Rodríguez, et al., 2014; Casanova, et al., 2018; Cejas, et al., 2019).

La evaluación posee esta importancia, por cuanto es la expresión del estado de logro estudiantil presente respecto al aprendizaje esperado -con base en el cual se toman decisiones pedagógicas- y por ello, es necesario reconocer que evaluar demanda entonces medir bien. Es decir, tener certeza sobre lo que se evalúa: Un requisito ineludible que hoy ya ha impulsado grandes avances desde el desarrollo de nuevos tipos de *test* y modelos de aplicación didácticos mediados por las tecnologías digitales (Rodríguez, et al., 2014).

Estas últimas han venido, justamente, adquiriendo un papel principal para ayudar a medir bien los aprendizajes, desde la mediación de los *test* informatizados. Estos últimos se refieren a evaluaciones generadas por un *software* especializado, que buscan medir y graduar en forma sencilla y rápida los conocimientos, habilidades y aptitudes que presenta un estudiante.

En los últimos años, ellos han dejado de ser una técnica experimental para constituirse en un procedimiento utilizado por cientos de programas de medición que evalúan a millones de personas cada año, hasta el punto de que se estima que hoy ocupan un lugar tan relevante como el que la evaluación escrita ha tenido hasta el presente (Rodríguez, et al., 2014; Ortiz-Colón, Ortega-Tudela y Román, 2019). En esta disposición es que, en dirección de no solo medir bien, sino, asimismo, de abordar en ámbitos educativos el problema evaluativo desde las tecnologías digitales, han surgido más recientemente los *Tests Adaptativos Informatizados* (TAI), que se consideran el siguiente nivel de las pruebas realizadas por computador (López, Pérez y Armendariz, 2005; Sierra-Matamoros, et al., 2007; López, 2008; López, Sanmartín y Méndez, 2014).

Estos *test*, a diferencia de los anteriores no siguen un patrón fijo. Es decir, que con ellos el estudiante no se adapta al instrumento, sino que, por el contrario, el instrumento se adapta al alumno, puesto que sus parámetros tienen como función amoldarse a las respuestas progresivas que va dando este último. El principal fundamento de los TAI es la Teoría de Respuesta al Ítem (TRI) (Muñiz, 2010; López, et al., 2014), que tiene como base la relación entre los valores que miden los *ítems* y la probabilidad de acertar en estos, denominada función de curva característica del *ítem* (CCI) (Muñiz, 2010). Para trabajar con base en esta función, los TAI se componen de propiedades psicométricas que permiten estimar, por ejemplo, la habilidad o rendimiento de una persona a partir de *ítems* (Olea y Ponsoda, 2002).

Es posible lograr, según Weiss (2004), si se asume una selección automática de *ítems* que funciona calculando una dificultad dinámica de selección, de acuerdo con la última respuesta obtenida de parte del usuario. De tal manera que, con los TAI, estos *ítems* son seleccionados automáticamente por un algoritmo,

hasta el momento en que se defina el nivel del evaluado (Olea y Ponsoda, 2002; Herrera, Barajas y Jiménez, 2015). En la actualidad, los TAI se han ido implementado como herramientas para identificar el nivel de competencias y habilidades de las personas (Ruiz, 2009). Pero, por contraste, su aplicación como instrumento educativo eficaz para la evaluación y toma de decisiones sobre los logros en el aprendizaje de la educación básica, no ha tenido igual desarrollo en países como Colombia (Herrera, et al., 2015).

En respuesta a tal necesidad, se ha planteado el propósito de analizar la aplicación de una prueba o *test* adaptativo informatizado (TAI) en educación básica, bajo el entendido de que ella permite ir avanzando en la construcción de conocimiento sobre propuestas claras para fortalecer y pedagogizar la evaluación en el ámbito escolar.

1. TEST ADAPTATIVO INFORMATIZADO (TAI)

1.1. Los TAI en el contexto internacional

En cuanto a los TAI en el ámbito internacional, se tiene para comenzar, Olea y Ponsoda (2002), quienes plantean que en EE.UU. desde 1980 se han realizado una serie de aplicaciones de exámenes adaptativos, como por ejemplo las pruebas CATASVAB o el GRE. Estos autores son los responsables de incorporar el concepto de *test* adaptativos al castellano, llevando este tipo de evaluación a la educación en España.

Por su parte, más adelante Abad, et al. (2002), lograron comparar la estimación de la habilidad, la precisión, el error y el sesgo en la aplicación de dos tipos de *tests* informatizados, los TAI, y los *tests* Óptimos (TO). Para ello, compararon las estimaciones a partir de tres métodos estadísticos descritos para TO (ML, MAP y EAP) y en relación a los TAI (ML), en una muestra de 88 estudiantes de primer curso de Psicología (con edades entre 17 y 19 años) de las Universidades Autónoma de Madrid y Santiago de Compostela. Cada uno respondió 20 *ítems* por *test*, desde un banco de 221 *ítems* de inglés. Encontraron que el TAI presentó el sesgo más bajo y no hallaron diferencias en el error típico.

En un lustro posterior, la investigación de Clares (2008), desarrolló y aplicó un TAI con el objetivo de establecer el menor número de preguntas y poder determinar el nivel de conocimiento de los estudiantes frente a una determinada materia. Para desplegarlo, tuvo en cuenta la construcción de un banco de preguntas, asumiendo lo propuesto por Olea y Ponsoda (1998). De esta manera, delimitó el nivel de los *ítems* en: Nivel bajo, medio y alto; y determinó la distribución de la dificultad dentro de cada nivel; asimismo, estableció la presentación del número y tipo de *ítems* al evaluado, para definir su nivel de conocimiento en un mínimo de dos fases. Como resultado, encontró dos formas para calcular el nivel de conocimiento, permitiendo dar orden en la presentación de los *ítems*, reducir el tiempo de aplicación y el número de estos, entre otros elementos relacionados con el evaluado.

Posteriormente, Escamilla, et al. (2014); y López, et al. (2014), hacen una revisión del modelo adaptativo de evaluación en educación para medir el progreso de los aprendizajes, estableciendo como antecedente que la evaluación adaptativa computarizada (CAT) o TAI (*test* adaptativo informático), fue sucesora de una serie de aplicaciones exitosas que surgieron a partir de 1905, con el desarrollo del primer examen adaptativo de Alfred Binet denominado *Binet IQ Test*.

Más recientemente, Toledo, et al. (2013), realizaron una contextualización alrededor de los TAI de mayor aplicación en la actualidad, entre las que se destacan: *Test of English as a Foreign Language* (TOEFL), el *Graduate Record Exam* (GRE) o el *Armed Services Vocational Aptitude Battery* (ASVAB), todos aplicados en EE.UU. Igualmente, lograron validar un modelo de evaluación adaptativa mediante razonamiento probabilista, haciendo uso de redes bayesianas, bajo el supuesto de la Teoría de Decisión Bayesiana (TDB). Así, se enfocaron en definir los elementos del perfil del estudiante al momento de construir un TAI y concluyeron que estas evaluaciones carecen de elementos esenciales del alumno tales como el perfil; por lo

que es necesario integrar a los algoritmos de evaluaciones adaptativos, variables que describen el perfil del estudiante.

Finalmente, Millán y Pérez (2014), establecieron un algoritmo a base de redes bayesianas para diagnosticar el conocimiento de contenidos en estudiantes, aplicando un TAI. Para validar el *test* adaptativo lo emplearon en una población simulada de estudiantes (en donde se conoce el 100% de los conceptos generados aleatoriamente). De esta forma, simulaban alumnos a partir de conjuntos preestablecidos de conceptos y niveles, logrando resultados correspondientes a aquellos criterios de selección que demostraron un mejor comportamiento para el diagnóstico y garantizando una calidad psicométrica del instrumento adaptativo en ambientes escolares. Así, aunque demostraron la aplicación del algoritmo de diagnóstico bayesiano, la misma se basó en simuladores y no directamente en estudiantes, disminuyendo así la capacidad explicativa del estudio.

En resumen, de acuerdo con la revisión sistemática de estos estudios, es posible reconocer que, si bien se ha venido progresivamente desarrollando el soporte tecnológico para aplicar los TAI en procesos escolares, es necesario incluir e integrar en ellos un diagnóstico de conocimientos hacia modelos más cognitivos, en donde se pueda evaluar, por ejemplo, el desempeño frente a un marco de competencias educativas en las que se forman los estudiantes.

1.2. Los TAI en el contexto colombiano

En relación con la aplicación de los TAI en Colombia, es necesario reseñar, en primer lugar, que Jiménez (2009) propuso la aplicación de un modelo de evaluación adaptativa exclusivamente desde el nivel de conocimientos, a través de un sistema tutorial inteligente. En un primer estudio de caso, utilizó 100 estudiantes simulados para un *test* de 20 preguntas y comparó la calificación entre un modelo tradicional y el adaptativo para una escala de conocimientos (muy bajo, bajo, medio, alto y muy alto), encontrando coincidencias en el nivel medio, debido a los parámetros de dificultad y discriminación.

En un segundo estudio de caso, se evaluó el estilo de aprendizaje, propuesto por Soloman y Felder (2005), para cuatro categorías, a 15 estudiantes de ingenierías, y se encontró que sólo un 30% logró un desempeño sobresaliente en tres de las categorías; los demás demostraron buen puntaje para una o dos categorías. En comparación con los resultados del *test* adaptativo, se estableció una relación directa entre dos de las categorías (Activo – Reflexivo y Visual - Verbal), concluyendo que uno de los mayores retos que tiene la implementación de la evaluación adaptativa en cursos virtuales, es la creación del banco de preguntas, puesto que este debe ser lo más completo posible; igualmente se encontró que la diferencia entre el modelo de evaluación tradicional y el adaptativo, está influenciado por los parámetros de las preguntas (dificultad) y la precisión.

Por otra parte, Herrera, et al. (2015) discuten la importancia de validar un TAI para valorar personas con limitaciones visuales o de baja visión, como una alternativa de evaluación para esta población. Con base en ello, realizan una revisión de los modelos de validación para los TAI, encontrando que los elementos para esta población: Garantizan una precisión en la magnitud de habilidad; posibilitan controlar algunas variables asociadas, como la participación de un lector (lo que afecta la validez de la prueba y el registro de información sobre el proceso de respuesta); brindan la posibilidad de hacer uso de nuevos formatos de *ítems*, proporcionando medidas más amplias de un dominio de constructo y dan mayor eficiencia en la medición de habilidades cognitivas de alto nivel.

Partiendo de dichos hallazgos, concluyen que la evaluación adaptativa permitiría superar algunas de las limitaciones que presentan las evaluaciones de lápiz y papel, indicando -sin embargo- que la adopción de estas pruebas adaptativas representa retos que deben valorarse adecuadamente, como lo son el uso de la tecnología en entornos de evaluación -especialmente cuando no son conocidos para los evaluados-; su percepción y los resultados ante las condiciones adaptativas, así como los costos económicos y exigencias tecnológicas para su implementación.

Un año más adelante, Restrepo, Barbosa y Arenas (2016) diseñaron y aplicaron una prueba adaptativa computarizada para la asignatura de Estática, con 118 estudiantes de los programas de Ingeniería Civil y Mecánica, en la Universidad EAFIT de Colombia. El estudio se suscitó debido al bajo rendimiento en la asignatura, cuyo índice de reprobación de estudiantes matriculados era de 48,8% y la deserción se encontraba alrededor del 30% en los registros oficiales. Para desarrollarlo, se implementó una plataforma virtual con el fin de mejorar los procesos de aprendizaje de los estudiantes y su rendimiento académico. Esta plataforma soportó una extensión para pruebas adaptativas.

Finalmente, el Instituto Colombiano para la Evaluación de la Educación (ICFES) inicia en el año 2018 una prueba piloto para aplicar una evaluación adaptativa, bajo el marco del proyecto “Piloto 4°, 6°, 8° Avancemos hacia un proceso formativo”, cuyo objetivo es convertirse en una nueva herramienta para fortalecer las prácticas docentes en el aula. A la fecha se han realizado dos aplicaciones en el país: La primera, se realizó en una muestra de 644.344 estudiantes de instituciones oficiales y privadas, para los componentes y competencias de Matemáticas y Lectura y escritura. El informe de resultados hizo énfasis en el porcentaje de preguntas correctas e incorrectas, relacionadas con las características de aprendizaje definidas en los parámetros de la prueba.

Por su parte, en la aplicación de 2019 de “Avancemos” participaron un total de 484.958 estudiantes de 4°, 6° y 8° grado de todo el país; de estos, el 32% fueron estudiantes de 8°, el 40% de 6° y el 28% de 4°. Hasta el momento, se ha desarrollado un aplicativo *web* que sintetiza los resultados, aunque estos solo están en coherencia con una pequeña muestra de los reguladores curriculares, lo cual limita sus intenciones frente a una muestra representativa del currículo.

En relación con las investigaciones y según Cortes, Castañeda y Daza (2019), es claro que a excepción de las pruebas desarrolladas recientemente por el ICFES en Colombia, se carece de experiencias en el uso de este tipo de *test* por limitantes en: El desarrollo y uso tecnológico para modular los TAI; el conocimiento técnico para su implementación; y, el proceso poco estructurado de las instituciones regulares de educación básica y media, para implementar modelos de evaluación formativa (Castro, Martínez y Figueroa, 2009); que además son exigidos desde el Decreto 1290 de 2009, por el cual se reglamenta la evaluación del aprendizaje y promoción de los estudiantes de los niveles de educación básica y media (Ministerio de Educación Nacional [MEN], 2009).

Con base en esto, es evidente que el país debe mejorar la calidad y coherencia de los procesos de evaluación interna de las instituciones, explorando modelos más innovadores en comparación con los tradicionales, como el que se busca aplicar en la presente investigación.

2. METODOLOGÍA

Con el propósito de implementar una aplicación de pruebas tipo TAI que aporte analíticamente al balance investigativo anteriormente examinado, la presente investigación partió de reconocer que la evaluación adaptativa, más allá de ser un proceso de calificación computarizado, es formativa en cuanto puede percibir cómo un ciclo de acción-reflexión sobre las preguntas fundamentales de la evaluación; se utiliza para obtener información detallada, que puede emplearse para modificar y mejorar el funcionamiento de un sistema o proceso. En este sentido, la evaluación desde este enfoque debería permitir adaptar el proceso individual de cada estudiante, detectar sus puntos débiles para poder corregirlos y tener un conocimiento completo y preciso de cada uno (Guzmán, 2005).

De lo anterior, que se conciba que un TAI es una prueba construida para fines de evaluación psicológica o educativa, cuyos *ítems* se presentan y responden mediante un ordenador, siendo su característica fundamental que se va adaptando al nivel de competencia progresivo que va manifestando el evaluado. Para cumplir este fin, los elementos básicos con que satisface la evaluación un TAI son, de acuerdo con Olea y Ponsoda (2002):

- a. Un banco de *ítems* con propiedades psicométricas conocidas, es decir, con parámetros estimados desde un modelo de la Teoría de la Respuesta al Ítem (TRI).
- b. Un procedimiento que establezca la manera de comenzar y finalizar la prueba, así como la forma de seleccionar progresivamente los mejores *ítems*.
- c. Un método estadístico de estimación de los niveles de rasgo o comportamiento.

Partiendo de estas concepciones y con el fin de realizar el análisis de la aplicación de la prueba adaptativa, se construyó un proceso de investigación correlacional basado en un diseño metodológico cuantitativo no-experimental de carácter transeccional, que permitiera dar cuenta de las posibles relaciones que la prueba podía tener al aplicarse sobre una muestra poblacional de estudiantes de ciclo III en colegios oficiales y privados de educación básica en Bogotá, Colombia (Hernández, Fernández y Baptista, 2014).

En este sentido, la investigación se desarrolló considerando dos momentos: Construcción y validación del instrumento. Para garantizar la calidad del instrumento y poder medir el comportamiento de los *test* adaptativos en la educación básica, se diseñó una prueba que lograra satisfacer las condiciones regulares de existencia del fenómeno evaluativo en Colombia, se determinó la validez de las preguntas seleccionadas del banco y confiabilidad del instrumento (Covacevich, 2014; Hernández, et al., 2014; Harrison, Collins y Müllensiefen, 2017, Castañeda, Camargo y Londoño, 2019, Daza, et al., 2019); asimismo, para su construcción se tuvo en cuenta el Modelo Basado en Evidencias (MBE), para definir las especificaciones de la prueba (ICFES, 2018).

En este caso, el instrumento se diseñó para un área de evaluación particular Ciencias Naturales (uno de los desafíos del MEN, 2004), considerando desde el banco de *ítems*, los siguientes parámetros: Componente, competencias, estándar, acciones de pensamiento, afirmaciones, evidencia, tarea y dificultad del *ítem*. El mismo consideró 19 *ítems* para ciclo III de educación básica, que fueron validados en su constructo y contaron con un pre-*test* cuya aplicación se realizó con una muestra promedio por *ítem* de 22.494 estudiantes de diferentes regiones del país, en el año 2018. Esta validación se realizó aplicando el α -Cronbach por medio del *software* estadístico SPSS que permitió determinar una confiabilidad de un 0,72 en la correlación de los *ítems* –incluida en el rango normal de confiabilidad reconocido.

Habiendo surtido esta validación de *ítems* de la prueba, se procedió entonces a incluirlos dentro de la estructura computacional de *Moodle* versión 3.6 que, aunada al *plugin* o extensión *Adaptive Quiz* (versión 1.2.3), permite conformar pruebas adaptativas alineadas a los requerimientos necesarios para este tipo de pruebas (Sierra-Matamoros, et al., 2007). La prueba estuvo estructurada por tres criterios: Uno de inicio (manifiesto en un *ítem* de dificultad baja por defecto); un criterio de selección de *ítems* (expresado en el aumento, disminución o mantenimiento de la dificultad del mismo de acuerdo con la última respuesta obtenida del estudiante); y un criterio de parada (que limitó a 7 *ítems* la prueba efectiva presentada por cada persona).

El algoritmo de la evaluación adaptativa en *Moodle* utilizó la técnica Rasch -con modelo logístico de un parámetro o IPL, asociado a la teoría de respuesta al *ítem*- teniendo como base de selección de éstos, los valores correspondientes al nivel de desempeño en que los mismos se ubicaban, del siguiente modo:

TABLA 1
Valores de selección y evaluación *ítems*

| Nivel de desempeño | Puntos de corte |
|--------------------|-----------------|
| Bajo | 1,0 - 1,66 |
| Medio | 1,67 - 2,3 |
| Alto | 2,31 - 3 |

Elaboración propia, 2020.

Presentados estos valores y considerando validado el diseño para la selección aleatoria de *ítems*, mediado por el algoritmo adaptativo informatizado de *Moodle* -cuyo error estándar estimado fue del 20%-, se consideró entonces realizada esta fase de la investigación y se procedió a desarrollar la selección de la muestra y aplicación de la prueba como segundo momento de ella.

Al respecto, se desarrolló el proceso de aplicación del instrumento validado anteriormente, sobre una muestra intencionada (Hernández, et al., 2014) de 544 estudiantes de educación básica pertenecientes a cinco colegios, tres oficiales y dos privados, de la ciudad de Bogotá en Colombia, distribuidos tal como se aprecia en la Tabla 2.

TABLA 2
Distribución de la muestra de estudiantes por sexo e institución

| Estudiantes | Sexo femenino | Sexo masculino |
|--------------------|---------------|----------------|
| Colegios Privados | 177 | 233 |
| Colegios Oficiales | 76 | 58 |
| Total | 253 | 291 |

Elaboración propia, 2020.

De igual manera, en la Tabla 3 se relacionan las características que se consideraron de estos colegios, para ser seleccionados como parte de la muestra intencionada.

TABLA 3
Características por instituciones vinculadas a la muestra

| Institución | Ubicación | Índole/Énfasis curricular | Clasificación de planteles ICFES (Saber 11) |
|-----------------------------------|------------------------------|--|---|
| Institución privada 1 | Localidad de Fontibón | Énfasis en media comercial con jornada única de preescolar a undécimo y una trayectoria de 63 años. | A+ |
| Institución privada 2 | Localidad de Teusaquillo | Énfasis en media académica con jornada única de preescolar a undécimo y una trayectoria de 82 años. | A+ |
| Institución privada 3 | Localidad Rafael Uribe Uribe | Énfasis en técnico comercial con jornada única. Presta servicio de preescolar a undécimo. | B |
| Institución Educativa Distrital 1 | Localidad de Engativá | Énfasis en media académica con triple jornada. Presta servicio de preescolar a undécimo y una trayectoria de 57 años. | A |
| Institución Educativa Distrital 2 | Localidad de La Candelaria | Énfasis en media técnica con doble jornada. Presta servicio de preescolar hasta grado undécimo y una trayectoria de más de 100 años. | C |

Entre los meses de junio y septiembre de 2019, se realizó la aplicación de la prueba adaptativa por computador, para lo cual se tuvo en cuenta la disponibilidad de ordenadores con acceso a *internet* por cada institución. La aplicación contó con un procedimiento que contempló un único intento por estudiante en un tiempo máximo de 25 minutos. Las pruebas no realizadas en su totalidad por parte de los alumnos, así como el reporte de más de un intento, fueron eliminados para la codificación de datos.

3. RESULTADOS Y DISCUSIÓN

El tratamiento de los datos obtenidos se realizó a través del paquete estadístico SPSS (Versión 24), aplicando la correlación de Pearson, para correlaciones entre variables cuantitativas; y la de Chi cuadrado, para las que integran una variable cualitativa y una cuantitativa. En primer lugar, para la muestra total de estudiantes observada, se obtuvo un puntaje promedio de 2,01 con una desviación estándar de 0,45, que corresponde al nivel de desempeño medio de todos los estudiantes evaluados en la prueba. En lo concerniente a la distribución general de puntajes por sexo, se encontraron resultados promedio idénticos de 2,01, diferenciados por la presencia de una desviación estándar mayor en el caso de las mujeres, cuyo comportamiento es posible inferir, sin embargo, por la cantidad de estudiantes evaluadas, que es menor con respecto al de alumnos hombres, tal como se puede apreciar en la Tabla 4.

TABLA 4
Distribución general de puntajes y desempeño por sexo para la muestra evaluada

| Sexo | No. Estudiantes | Porcentaje de participación (%) | Puntaje | Desviación Estándar | Desempeño |
|-----------|-----------------|---------------------------------|---------|---------------------|-----------|
| Masculino | 291 | 53,4 | 2,01 | 0,435 | Medio |
| Femenino | 253 | 46,6 | 2,01 | 0,477 | Medio |

Elaboración propia, 2020.

Por su parte, en el caso de la relación por tipo de institución, se encontró que, a pesar de obtener un desempeño “Medio”, tanto en instituciones oficiales como privadas, existe una diferencia entre los puntajes, puesto que, aunque las privadas presentan el mayor porcentaje de participación, se encuentran al menos un 0,4 por arriba de las oficiales, incluyendo un porcentaje de desviación estándar menor que en estas últimas, como se muestra en la Tabla 5.

TABLA 5
Distribución general de puntajes y desempeño por tipo de institución para la muestra evaluada

| Tipo de institución | No. Estudiantes | Porcentaje de participación (%) | Puntaje | Desviación Estándar | Desempeño |
|---------------------|-----------------|---------------------------------|---------|---------------------|-----------|
| Privada | 410 | 75,4 | 2,15 | 0,39 | Medio |
| Oficial | 134 | 24,6 | 1,65 | 0,43 | Medio |

Elaboración propia, 2020.

En relación a las instituciones participantes, los puntajes más altos se encuentran asociados a los colegios de carácter privado; asimismo, frente a los desempeños, sólo una de las instituciones se ubicó en desempeño “Bajo”, tal como se evidencia en la Tabla 6.

TABLA 6
Distribución de puntajes y desempeño por institución educativa

| Institución Educativa | No. Estudiantes | Porcentaje de participación (%) | Puntaje | Desviación Estándar | Desempeño |
|-----------------------------------|-----------------|---------------------------------|---------|---------------------|-----------|
| Institución Privada 1 | 171 | 31,4 | 2,12 | 0,40 | Medio |
| Institución Privada 2 | 189 | 34,7 | 2,18 | 0,38 | Medio |
| Institución Educativa Distrital 1 | 50 | 9,2 | 1,98 | 0,37 | Medio |
| Institución Educativa Distrital 2 | 28 | 5,1 | 1,86 | 0,38 | Medio |
| Institución Educativa Distrital 3 | 106 | 19,5 | 1,59 | 0,42 | Bajo |

Elaboración propia, 2020.

Finalmente, respecto a las variables cualitativas, no hay diferencias significativas entre el desempeño en la prueba y el sexo de los participantes. Sin embargo, se evidencian diferencias relevantes entre el desempeño y el carácter de la institución (oficial y privado). En la correlación de las variables “Desempeño” y “Sexo”, se encontró que no existen diferencias significativas que permitan asociarlas. En cambio, en el caso de las correlaciones entre “Desempeño” y “Carácter de la institución” y “Desempeño” e “Institución”, los resultados arrojaron un porcentaje altamente significativo, tal como se puede evidenciar en la Tabla 7.

TABLA 7
Coeficientes de correlación entre desempeños y variable sexo, carácter de la institución e institución

| Relación | Características | Valor | gl | Significancia |
|-----------------------|-----------------|--------------------|----|---------------|
| Desempeño/Sexo | χ^2 | 0,846 ^a | 2 | 0,655 |
| | Casos válidos | 544 | | |
| Desempeño/Carácter | χ^2 | 94,343 | 2 | 0,000 |
| | Casos válidos | 544 | | |
| Desempeño/Institución | χ^2 | 125,153 | 8 | 0,000 |
| | Casos válidos | 544 | | |

Elaboración propia, 2020.

Estos resultados permiten observar cómo a partir de un instrumento de evaluación con adaptabilidad, resulta potencial el análisis de variables asociadas al desempeño, que van más allá del valor cuantitativo en que se transforma la evaluación y permiten dar cuenta de posibles factores influyentes en los procesos de enseñanza-aprendizaje evaluados en el instrumento.

Así, por ejemplo, el que las instituciones de carácter privado presenten en promedio un mayor desempeño que las de carácter público, aun contando con una mayor variabilidad muestral por el número de estudiantes

evaluados, puede llegar a estar asociado, como se ha reconocido en investigaciones recientes (Vasco, 2015), a diferentes factores como son:

1. Los socioeconómicos, en donde se ha identificado que el nivel de ingresos y el capital cultural de estudiantes en instituciones de carácter privado en Bogotá, puede llegar a ser mayor que el de las de carácter público;
2. Los de gestión, en donde factores como la estabilidad docente, el clima institucional y el mayor acompañamiento de los padres, pueden llegar a influir en la relación que tienen los estudiantes con su educación y los resultados en calidad que de ello resulta;
3. Los pedagógicos y evaluativos, en la medida en que la exigencia sobre los procesos pedagógicos que guía el profesorado, puede llegar a tener una mayor claridad y capacitación.

Teniendo en cuenta estos factores, sería posible a su vez entender la diferencia observada entre las instituciones por separado que, igualmente, muestran que existe una tendencia de las instituciones de carácter privado a obtener mayores desempeños cuantitativos, aunque se encuentren en rangos cualitativos de desempeño semejantes a los de las instituciones públicas.

Más, si esto sucede en el análisis particular de posibles factores asociados al desempeño en cada institución, igualmente puede pasar con la efectividad del análisis que provee la prueba adaptativa, para establecer comparaciones fiables entre grupos de estudiantes e instituciones. Si bien es posible que no todos los estudiantes hayan contestado la misma secuencia de ítems incluidos dentro de la prueba, la exactitud de esta coincidencia no termina por afectar el resultado, en la medida en que lo evaluado da cuenta de competencias articuladas a los reguladores del currículo, que viabilizan la realización de dicha comparación.

Asimismo, en la correlación del desempeño de los estudiantes en la prueba y la institución, se encontró un alto nivel de significancia entre las variables ($p = 0,001$), esto indica un alto nivel de dependencia y por tanto una fuerte relación entre estas. Teniendo en cuenta la caracterización de las instituciones educativas participantes en el estudio, este resultado puede estar afectado por el énfasis de la formación y la clasificación en las pruebas de estado Saber (ICFES, 2013; 2014 y 2018), que es un indicador de la calidad educativa que se imparte al interior de cada una de ellas.

De esta forma, la correlación entre los desempeños obtenidos y la variable “Sexo”, a pesar de ser mayor el número de estudiantes hombres evaluados, la diferencia de desempeños con respecto al de las mujeres no es significativa. Este resultado, es similar a lo encontrado por Correa (2016), quien comparó el desempeño académico de los estudiantes y su diferencia en cuanto género (hombre y mujeres), entre los grados cuarto a octavo del sistema educativo colombiano y no encontró diferencias entre los grupos de comparación, aunque se reconoce que en condiciones de brechas entre los géneros, algunos factores como el nivel de autoconfianza puede afectar a las mujeres en su desempeño académico.

No obstante, frente al desempeño de los estudiantes en la prueba y el “Carácter de la institución” (oficial y privado), se evidencia un alto nivel de significancia entre las variables ($p = 0,001$), lo que indica que, el carácter de la institución afecta el desempeño de los estudiantes en la prueba, y teniendo en cuenta que las instituciones privadas presentan un mejor desempeño, esto puede deberse a que el carácter de la institución puede estar relacionado con algunos factores socioeconómicos, como la renta media del hogar, el nivel educativo de los padres y su ocupación, mayores recursos y oportunidades para ellos (Correa, 2016).

En el presente estudio no se evaluaron factores socioeconómicos, estos podrían explicar el por qué de las diferencias entre el carácter de la institución (oficial y privado). Así, es posible reconocer que su relevancia en el marco del enfoque evaluativo que se asumió pueda llegar a ser menor, en la medida en que las comparaciones a que da lugar el TAI tienen por propósito comparar posibles factores asociados a desempeños y adaptabilidades semejantes y no establecer comparaciones sin tener en cuenta el nivel de desempeño real del estudiante, como puede pasar con evaluaciones tradicionales.

CONCLUSIONES

Frente a la aplicación de la prueba adaptativa tipo TAI con los estudiantes del ciclo III de educación básica de las instituciones participantes, se pudo evidenciar el desempeño alcanzado según los aprendizajes esperados. En lo particular, las instituciones de carácter privado presentan en promedio un mayor desempeño que las de carácter público, aun contando con una mayor variabilidad muestral por el número de estudiantes evaluados, lo que podría estar asociado con algunos factores socioeconómicos, de gestión académica, pedagógicos y evaluativo.

En lo relacionado al proceso evaluativo la aplicación de los TAI en Instituciones Educativas de Colombia, pueden ayudar a fortalecer componentes de la evaluación escolar y el progreso de los estudiantes frente a los aprendizajes esperados para el proceso, si se enfocan en desempeños de competencia. Asimismo, los TAI se convierten en una alternativa a los test tradicionales de papel y lápiz, en la medida en que, además de ser más eficientes, permiten que el instrumento se adapte al evaluado y no lo contrario, visibilizando aquellas características o rasgos que requieren ser mejorados en el proceso (Orjuela, et al., 2020).

Los resultados obtenidos en la prueba aplicada en el estudio, permitió formular las siguientes recomendaciones, si las instituciones educativas quieren que se mejoren y fortalezcan los procesos de evaluación implementando los TAI o pruebas adaptativas, como: Realizar una evaluación frente a los requerimientos tecnológicos que se necesitan al interior de la institución, puesto que esto requiere como mínimo la instalación de una plataforma tipo LMS o software, la disponibilidad de equipos de computos para las aplicación de las pruebas, un servidor para el almacenamiento de datos y personal cualificado, que pueda brindar el soporte técnico a la hora de gestionar los usuarios y las pruebas.

De igual manera, se requiere de una conceptualización pedagógica de la “Evaluación Adaptativa” por parte del equipo docente, con el fin de lograr comprender su alcance y limitaciones al interior de la evaluación de los aprendizajes, lo cual se vuelve crucial debido a que es un tema novedoso y poco conocido en las instituciones.

Finalmente, el grado de dominio o nivel de desarrollo de las competencias tecnológicas que se requiere por parte del equipo docente, es fundamental, puesto que no sólo se requiere de una exploración e implementación de una herramienta tecnológica para apoyar los procesos de evaluación, sino que es necesario llegar al nivel de innovación, es decir, hacer parte del desarrollo y concepción de la evaluación adaptativa en coherencia con el contexto educativo así como los aprendizajes esperados para el proceso de formación.

REFERENCIAS BIBLIOGRÁFICAS

- Abad, F. J., Olea, J., Real, E. y Ponsoda, V. (2002). Estimación de habilidad y precisión en test adaptativos informatizados y test óptimos: Un caso práctico. *Revista Electrónica de Metodología Aplicada*, 7(1), 1-20.
- Casanova, I., Canquiz, L., Paredes, Í., e Inciarte, A. (2018). Visión general del enfoque por competencias en Latinoamérica. *Revista de Ciencias Sociales (Ve)*, XXIV(4), 114-125.
- Castañeda, J. G., Camargo, J. A., y Londoño, M. K. (2019). Análisis bibliométrico como herramienta para el seguimiento de instrumentos psicológicos validados en Colombia. *Revista Ibérica de Sistemas e Tecnologías de Informação*, (E-18), 38-46.
- Castro, H., Martínez, E., y Figueroa, Y. (2009). *Fundamentaciones y orientaciones para la implementación del Decreto 1290 del 16 de abril de 2009. Documento No. 11*. Ministerio de Educación Nacional. https://www.mineducacion.gov.co/1759/articles-213769_archivo_pdf_evaluacion.pdf
- Cejas, M. F., Rueda, M. J., Cayo, L. E., y Villa, L. C. (2019). Formación por competencias: Reto de la educación superior. *Revista de Ciencias Sociales (Ve)*, XXV(1), 94-101.
- Celman, S. (2005). ¿Es posible mejorar la evaluación y transformarla en una herramienta de conocimiento? En A. R. de Camilloni (Ed.), *La evaluación de los aprendizajes en el debate didáctico contemporáneo* (pp. 35-66). Paidós Educador.

- Clares, J. (2008). Propuesta de desarrollo de test informatizado adaptándolo a las respuestas del usuario. *Pixel-Bit: Revista de Medios y Educación*, (31), 19-30.
- Correa, J. (2016). Desempeño académico y diferencias de género en Colombia: Un análisis con base en las pruebas TIMSS 2007. *Sociedad y Economía*, (30), 15-42.
- Cortes, J. E., Castañeda, J. G., y Daza, J. (2019). Comprensión lectora de estudiantes universitarios. Factores asociados y mecanismos de acción. *Revista Venezolana de Gerencia*, 24(87), 874-889.
- Cortes, J. E., Daza, J., y Castañeda, J. G. (2019). Relación del entorno socioeconómico con el desempeño de la comprensión lectora en universitarios. *Revista de Ciencias Sociales (Ve)*, XXV(4), 119-133.
- Covacevich, C. (2014). *Cómo seleccionar un instrumento para evaluar aprendizajes estudiantiles*. Banco Interamericano de Desarrollo BID. Recuperado de <https://publications.iadb.org/publications/spanish/document/C%C3%B3mo-seleccionar-un-instrumento-para-evaluar-aprendizajes-estudiantiles.pdf>
- Daza, J., Castañeda, J. G., Torres, C., Segovia, C., y Cortes, E. (2019). Diseño y análisis psicométrico de una prueba para medir la percepción de clases frente a la formación integral de estudiantes universitarios «PCFI». *Revista Espacios*, 40(2), 18-29.
- Escamilla, J., Calleja, B., Villalbam, E., Venegas, E., Fuerte, K., Román, R., y Madrigal, Z. (2014). *Aprendizaje y evaluación adaptativos*. Reporte Edu Trends. Observatorio de Innovación Educativa del Tecnológico de Monterrey. <https://observatorio.tec.mx/edutrendsaprendizajeadaptativo>
- Guzmán, D. (2005). *Un modelo de evaluación cognitiva basado en Tests Adaptativos para el diagnóstico en Sistemas Tutores Inteligentes* (Tesis doctoral). Universidad de Málaga, España.
- Harrison, P. M. C., Collins, T., y Müllensiefen, D. (2017). Applying modern psychometric techniques to melodic discrimination testing: Item response theory, computerised adaptive testing, and automatic item generation. *Scientific Reports*, 7, 3618. <https://doi.org/10.1038/s41598-017-03586-z>
- Hernández, S., Fernández, C., y Baptista, L. (2014). *Metodología de la investigación*. McGraw-Hill.
- Herrera, A. N., Barajas, R., y Jiménez, G. J. (2015). Validez en test adaptativos informatizados: Alternativa para evaluar población con limitaciones visuales. *Avaliação Psicológica*, 14(3), 299-307.
- Instituto Colombiano para la Evaluación de la Educación - ICFES (2013). *Sistema Nacional de Evaluación Estandarizada de la Educación Alineación del examen SABER 11°*. <https://www.icfes.gov.co/documents/2014/3/193784/Alineacion%20examen%20Saber%2011.pdf>
- Instituto Colombiano para la Evaluación de la Educación - ICFES (2014). *Pruebas Saber 3°, 5° y 9° Lineamientos para las aplicaciones muestral y censal 2014*. <https://www.icfes.gov.co/documents/2014/3/176813/Guia+de+lineamientos+para+las+aplicaciones+muestral+y+censal+-+saber+359+2014.pdf/7455a026-8308-5d4d-4975-a53a86e4c56b>
- Instituto Colombiano para la Evaluación de la Educación - ICFES (2018). *Avancemos .°, 6° y 8°: Hacia un proceso formativo*. Boletín Saber en Breve. Edición 31. <http://www.icfes.gov.co/edicion-31-boletin-saber-en-breve>
- Jiménez, M. (2009). *Un modelo de evaluación adaptativa del nivel de conocimientos en Sistemas Tutoriales Inteligentes* (Tesis de maestría). Universidad Nacional de Colombia, Colombia.
- López, C., Pérez, T., y Armendariz, A. (2005). Evaluación mediante tests: ¿Por qué no usar el ordenador? *Revista Iberoamericana de Educación*, 36(11), 1-16.
- López, J. (2008). *Evaluación mediante test adaptativos informatizados en el contexto de un sistema adaptativo para el aprendizaje de la lengua vasca* (Tesis doctoral). Universidad del País Vasco, San Sebastián, España.
- López, R., Sanmartín, P., y Méndez, F. (2014). Revisión de las evaluaciones adaptativas computarizadas (CAT). *Educación y Humanismo*, 16(26), 27-40.
- Millán, E., y Pérez, J. L. (2014). Un algoritmo de diagnóstico para modelado del alumno basado en test adaptativos y redes bayesianas. <http://www.lcc.uma.es/~eva/doc/materiales/millane.pdf>
- Ministerio de Educación Nacional - MEN (2004). *Formar en Ciencias: ¡El desafío! Estándares básicos de competencias en Ciencias Naturales y Ciencias Sociales*. Serie guías No. 7. https://www.mineducacion.gov.co/1759/articles-81033_archivo_pdf.pdf

- Ministerio de Educación Nacional - MEN (2009). *Decreto 1290. Por el cual se reglamenta la evaluación del aprendizaje y promoción de los estudiantes de los niveles de educación básica y media*. https://www.mineducacion.gov.co/1621/articles-187765_archivo_pdf_decreto_1290.pdf
- Muñiz, J. (2010). Las teorías de los tests: Teoría Clásica y Teoría de Respuesta a los ítems. *Papeles del Psicólogo*, 31(1), 57-66.
- Olea, J., y Ponsoda, V. (1998). Tests informatizados y adaptativos informatizados: Investigación en España. *Relieve*, 4(2), 1-3.
- Olea, J., y Ponsoda, V. (2002). Test adaptativos informatizados. En J. Muñiz (Coord.), *Psicometría* (pp. 729-784). Universitat.
- Orjuela, C. H., Araque, G. A., Lesmes, O. M., y Castañeda, J. G. (2020). Estrategias de aprendizaje adaptativo en educación y ciencias sociales: pruebas adaptativas computarizadas. *Revista Venezolana de Gerencia*, 25(E-3), 482-492.
- Ortiz-Colón, A. M., Ortega-Tudela, J. M., y Román, S. (2019). Percepciones del profesorado ante la alfabetización mediática. *Revista de Ciencias Sociales (Ve)*, XXV(1), 11-20.
- Restrepo, J. L., Barbosa, J. L., y Arenas, J. (2016). Aplicación de una prueba adaptativa computarizada en la asignatura Estática. *14th LACCEI International Multi-Conference for Engineering, Education, and Technology: "Engineering Innovations for Global Sustainability"*, San José, Costa Rica.
- Rodríguez, C., Muñoz, M., y Castillo, V. (2014). Test informatizados y su contribución a la acción evaluativa en educación. *RED. Revista de Educación a Distancia*, (43), 136-152.
- Ruiz, M. (2009). Los tests adaptativos informatizados y su utilidad en el contexto clínico. *Conexionismo*. http://www.conexionismo.com/leer_articulo.php?ref=los_tests_adaptativos_informatizados_y_su_utilidad_en_el_contexto_clinico-2picnex0
- Shepard, L. A. (2006). La evaluación en el aula. En R. L. Brennan (ed.), *Educational Measurement* (pp. 623-646). Praeger Publishers.
- Sierra-Matamoros, F. A., Valdelamar-Jiménez, J. R., Hernández-Tamayo, F. A., y Sarmiento-García, L. M. (2007). Test adaptativos informatizados. *Avances en Medición*, 5(1), 157-162.
- Soloman, B., Felder, R. (2005). Index of Learning Styles Questionnaire. NC State University. <http://www4.ncsu.edu/unity/lockers/users/f/felder/public/ILSpage.html>
- Toledo, G., Mezura, C., Cruz, N., y Benítez, E. (2013). Modelo de evaluación adaptativa y personalizada mediante razonamiento probabilista. <https://docplayer.es/16128419-Modelo-de-evaluacion-adaptativa-y-personalizada-mediante-razonamiento-probabilista.html>
- Vasco, C. (2015). Diez retos de la educación colombiana para 2025. En J. Zubiría (Comp.), *La calidad de la educación bajo la lupa* (pp. 23-32). Editorial Magisterio. 246 pp.
- Weiss, D. J. (2004). Computerized adaptive testing for effective and efficient measurement in counseling and education. *Measurement and Evaluation in Counseling and Development*, 37(2), 70-84.
- Yovane, K. (2000). La evaluación de los aprendizajes en la reforma curricular de la Educación Básica. *Acción Pedagógica*, 9(1y2), 12-16.