

Diseño y validación de pruebas diagnóstico en el área de tecnología para el ciclo V de secundaria

Vargas, Luis Fernando, Olarte, Fredy Andres, Ramirez, Jhon Jairo

Resumen

El propósito del presente artículo es describir el proceso de diseño y validación de dos pruebas diagnósticas, para evaluar un conjunto de competencias en educación tecnológica, en los niveles de décimo y undécimo. Se inicia con un breve marco conceptual y contextual acerca de las competencias en tecnología y su organización en el contexto colombiano a través del documento Serie Guía # 30 (MEN). El documento no evidencia herramientas evaluativas, por ello se construyeron dos pruebas de tipo diagnóstico para medir las competencias tecnológicas en el estudiante. La validación y fiabilidad incluyó un proceso de pilotaje con estudiantes de la media, de dos instituciones del distrito capital. Los resultados obtenidos permiten establecer, índice de confiabilidad adecuados, lo mismo que índices de discriminación y dificultad en cada una de las pruebas, hallando con ello el nivel que tienen los estudiantes en las competencias en tecnología.

Palabras Clave: Educación en tecnología, validación, confiabilidad, competencias.

I. INTRODUCCIÓN

La formación por competencias para el desarrollo de aprendizajes se ha configurado como una estrategia educativa extendida a nivel mundial [1]. A pesar de las virtudes de dicho modelo, la evaluación de estas se presenta como un reto importante dentro del proceso formativo de los estudiantes. Este reto implica reconocer en los estudiantes la capacidad de aplicar habilidades y conocimientos frente a un problema que pertenece a un entorno determinado [2]. El proceso evaluativo esta compuesto de tres partes fundamentales la evaluación diagnóstica, la formativa y la sumativa. La diagnóstica se desarrolla al inicio del proceso y permite determinar aquellos preconceptos que el estudiante tiene antes de iniciar el curso. La formativa se desarrolla a lo largo del proceso educativo y le permite al docente establecer mediante diversos mecanismos, el nivel que el estudiante tiene en cada una de las competencias trabajadas. La sumativa permite determinar al final de un periodo académico los alcances que el estudiante obtuvo [3]. La evaluación diagnóstica y sumativa hacen medición sistemática y gradual, por tanto entregan como resultado el nivel de logro obtenido por el estudiante [4].

Determinar el nivel que tiene el estudiante en cada una de las etapas de la evaluación, implica tener instrumentos de medición validados [5]. La validación es un aspecto importante, debido a que un instrumento con índices de

validez bajos, probablemente puede reportar información incorrecta sobre el desempeño del estudiante. Es por ello que al momento en que se crea un instrumento evaluativo es necesario indagar acerca de su capacidad de medir la competencia, para la cual se creo [6].

Experiencias como las de [1], [5] y [7] presentan la necesidad de validar los instrumentos evaluativos, ya que ello contribuye a mejorar en el proceso educativo. La primera de ellas plantea que las pruebas escritas bien desarrolladas y validadas de forma adecuada, brindan información que permite determinar el nivel de competencia que tiene el estudiante. En el segundo caso se presenta una mirada a los diversos insumos que se requieren para el diseño de pruebas escritas y su posterior validación. Inicia con la competencia, sus dimensiones o desempeños y estos medidos mediante indicadores.

En el caso de la educación tecnológica se tienen diversas experiencias que trabajan mediante competencias. Trabajos como los de [8], [9] y [10], evidencian la necesidad de trabajar con mecanismos evaluativos que den información sobre el proceso educativo del estudiante. Estos mecanismos deben evidenciar los avances obtenidos por el estudiante a través del proceso, deben brindar información acerca de las habilidades en las que tienen dificultades y en aquellas donde tienen fortalezas. Para poder determinar el avance desde el inicio al final del proceso es necesario crear instrumentos diagnósticos. El mas común de ellos es las pruebas, las cuales han sido usadas para determinar el estado inicial del estudiante. En Colombia si bien se tienen las competencias tecnológicas, no se tienen instrumentos que permitan determinar el nivel inicial de los estudiantes. La normatividad establecida por el MEN no presenta indicadores que permitan medir los niveles de logro obtenidos por los estudiantes, por tanto no se establecen los conceptos previos, ni las habilidades tecnológicas que tienen los estudiantes antes de iniciar el curso. Es por ello que en el presente trabajo se presentan las etapas seguidas para el diseño y validación de dos pruebas diagnósticas.

A continuación, se describe la organización dada por el MEN en la educación tecnológica. En la siguiente sección presenta el proceso de diseño de las pruebas diagnóstico, así como las características que definen el proceso de validación y confiabilidad de instrumentos evaluativos. En la cuarta sección del documento, se describe el proceso propuesto para desarrollar la aplicación del piloto en las dos instituciones distritales. En la quinta sección se analizan los resultados obtenidos en el proceso de validación considerando variables como índice de dificultad, de discriminación y confiabilidad de cada prueba. Por último, en la sexta sección se presentan las conclusiones obtenidas y los trabajos futuros a desarrollar.

Primer Autor: lufvargasne@unal.edu.co, estudiante de maestría en automatización industrial. Universidad Nacional de Colombia.

Segundo Autor: faolarted@unal.edu.co, Profesor, Departamento de Ingeniería, Universidad Nacional de Colombia.

Tercer Autor: jjramireze@unal.edu.co, Profesor, Departamento de Ingeniería, Universidad Nacional de Colombia.

II. ORGANIZACIÓN DE LA EDUCACIÓN TECNOLÓGICA EN COLOMBIA

La educación en tecnología es reconocida internacionalmente por ser un área fundamental, que permite a los futuros ciudadanos participar activamente en la sociedad [11]. En Colombia, el Ministerio de Educación Nacional (MEN) estableció las directrices para la enseñanza de la tecnología e informática a través de un documento donde se encuentran las competencias y desempeños que los estudiantes de primaria y secundaria deben alcanzar. Estas competencias deben ser desarrolladas en los planes de estudio de las diferentes instituciones de básica y media, impartida de forma obligatoria en cada una de ellas [12].

El MEN en el año 2008 publicó el documento '*serie 30: ser competente en tecnología: una necesidad para el desarrollo*' en él se da una organización a los aspectos relevantes de la educación en tecnología. El documento inicia dando un marco conceptual en el cual se establecen algunas definiciones básicas alrededor del significado de la tecnología. Posteriormente presenta las relaciones que tiene la tecnología con la evolución del hombre, la ciencia, la técnica, la innovación, la ética, el desarrollo y evolución de la sociedad. Finalmente plantea un grupo de lineamientos, con los cuales es posible organizar los procesos educativos y evaluativos en el aula, mediante una jerarquía de componentes, competencias y desempeños para cada uno de los grupos de grados de educación básica y media [12].

La jerarquía presentada en la serie 30 presenta cuatro componentes en su primer nivel: *naturaleza y evolución de la tecnología, apropiación y uso de la tecnología, solución de problemas con tecnología y tecnología y sociedad*. El segundo nivel de cada componente presenta la competencia que se espera desarrollar en el estudiante en su proceso educativo. Por cada competencia se tienen un conjunto de desempeños o dimensiones con el fin de obtener una medición del alcance de la competencia por parte del estudiante.

No obstante, la serie 30 no presenta en su organización los indicadores de desempeño que se deben desarrollar en el proceso de aprendizaje. Esto implica que no se tienen mecanismos que permitan establecer mediciones acerca de las habilidades que se deben potenciar. Por esto los docentes no suelen evaluar las competencias, sino los conocimientos impartidos sobre alguna temática. De igual manera no se establecen los conceptos previos, ni las habilidades tecnológicas que tienen los estudiantes antes de iniciar el curso. Por ende es necesario crear un instrumento diagnóstico que contribuya al docente y a los estudiantes en establecer el nivel previo de las competencias tecnológicas. Por ello se crearon y validaron dos pruebas diagnósticas para dos de los componentes, *solución de problemas con tecnología y apropiación y uso de la tecnología*.

III. DISEÑO Y VALIDACIÓN DE INSTRUMENTOS EDUCATIVOS

La validación de un instrumento educativo brinda la información necesaria acerca de la veracidad de los datos obtenidos por el instrumento creado [7]. En la figura 1 se

resume las fases seguidas para obtener datos que permitan validar las pruebas diagnósticas creadas. La figura permite observar que se tienen dos etapas principales, el diseño de pruebas y la validación de éstas. Dentro de cada etapa se tiene una serie de fases que permiten al ser completadas obtener un insumo. En el primer caso al terminar las fases de conceptualización y construcción de reactivos se obtendrán las pruebas diagnósticas. En el segundo caso al finalizar las fases internas se obtendrá la validación de las pruebas.

Cada una de estas etapas serán descritas en las siguientes secciones, en la primera de ellas se explica lo relacionado con el diseño de las pruebas, a partir de los tipos de preguntas usados y su selección. En la segunda etapa se presentan los tipos de validez usados y la confiabilidad, para cada una de las pruebas diagnósticas construidas.

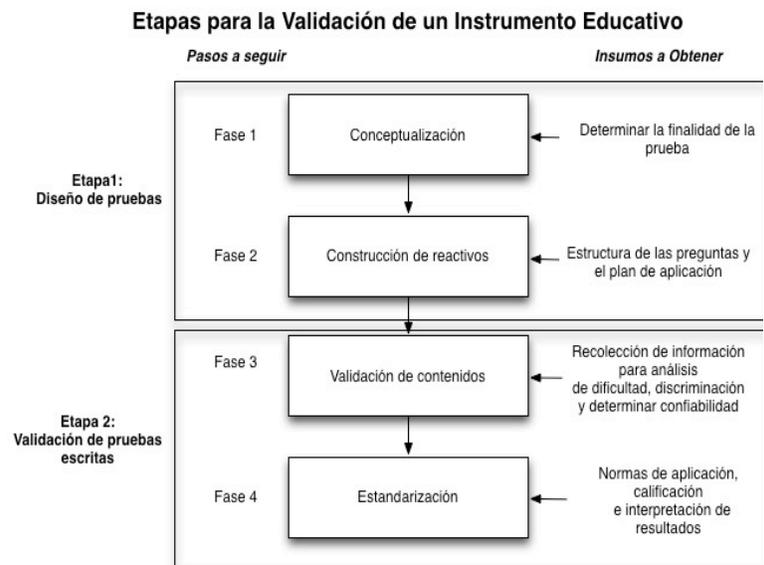


Fig.1. Fases utilizadas para la validación de un instrumento educativo [7]

A. Diseño de pruebas diagnósticas

La evaluación es un proceso que debe ser planeado y establecido a partir de un conjunto de metas educativas claras [13]. En el caso de la educación en tecnología en Colombia los objetivos de aprendizaje no están del todo claros. Si bien se definen los desempeños y competencias [12], no se tienen indicadores, a partir de los cuales generar herramientas para desarrollar el proceso evaluativo. Por ello se hizo necesario construir los indicadores de desempeño, ya que estos permiten determinar los pasos a seguir en el proceso evaluativo [14].

El primer paso en el diseño de las pruebas fue traducir los desempeños en habilidades para determinar el nivel cognitivo al que pertenecen, de acuerdo con la taxonomía de Bloom revisada [15]. Se tomó el verbo principal de cada desempeño, se categorizó en alguno de los dominios cognitivos dentro de la escala de Krathwohl y se concluyó el alcance de aprendizaje, la cualidad que desarrolla, así como la posible evidencia a obtener en el estudiante [16]. Los aspectos restantes del desempeño, permiten determinar los alcances esperados en el alumno. De esta manera se establecen los

indicadores que permitirán obtener evidencia de los aprendizajes que tiene u obtendrá el estudiante [17].

Al tener claridad acerca de los alcances de cada desempeño y las habilidades cognitivas que se esperan, se inició con la construcción de las pruebas. Para ello se tomaron en cuenta algunos aspectos de las pruebas PISA, en particular el que establece que una prueba escrita no debe sólo determinar conocimientos, habilidades, y aptitudes del estudiante, sino que además debe evidenciar la aplicación de estas características en diversos entornos [18]. De esta manera se tomó cada una de las habilidades y se tradujo en una serie de preguntas que tienen en común el solucionar problemas dentro de un contexto.

Para la construcción de las preguntas se hizo necesario establecer los tipos de preguntas posibles, de acuerdo con las habilidades a revisar. Se exploró con preguntas de verdadero y falso, de apareamiento, de completar entre otras, pero muchas de ellas no satisfacían las habilidades a medir. Por tanto, se estableció que las preguntas de tipo abierto y cerrado con respuesta única y múltiple, podrían ser adecuadas para este caso. Estos tipos de pregunta permiten determinar el nivel de competencia de un estudiante de forma adecuada [19].

En el caso de las preguntas abiertas, estas le brindan al estudiante la posibilidad de mostrar sus habilidades y evidencian los niveles más altos que se tienen en la taxonomía de Bloom. Para su construcción se siguió la guía dada por Husain, Bais, Hussain y Samadb [20]. En ella se plantean cinco criterios para su construcción:

- *Concepto en un campo determinado:* los elementos de la evaluación transmitidos al estudiante, deben recalcar lo importante. Dar una serie de objetivos con los cuales el estudiante tenga la posibilidad de comprender mediante vinculación todo el tema y con él, solucionar problemas.
- *Una pregunta puede tener varias respuestas:* la pregunta debe evidenciar que se tienen varias formas de solución. Esto implica usar los conceptos de la posible respuesta de diversas maneras, lo que genera formas de pensamiento distintas del problema para cada estudiante. Contribuyendo a la argumentación, al analizar y “criticar” al momento de solucionar un problema en un contexto.
- *Necesidad de comunicar el proceso de razonamiento:* dar al estudiante la posibilidad de comunicar las ideas que tienen en su mente, acerca de un tema determinado. Esto genera en el estudiante la necesidad de no solo argumentar, sino de enlazar distintos conocimientos.
- *Las preguntas deben quedar claramente establecidas:* el estudiante al momento de leer la pregunta debe tener claridad en lo que se espera de él y cómo debe ser la respuesta a la pregunta. Este aspecto es importante ya que permite desarrollar mejora en sus habilidades comunicativas, de pensamiento analítico y crítico.
- *Tener un modelo de calificación:* cada pregunta a evaluar debe tener por lo menos una matriz de

valoración de dos puntos Sí/No. Pero la idea de las preguntas abiertas es generar en el estudiante la posibilidad de comunicar sus conocimientos en un contexto determinado. Por tanto es necesario tener un sistema de respuestas basadas en las posibilidades que se pueden tener. Para ello es necesario conocer que una buena pregunta, tiene asociada una rúbrica, como indicativo del valor de la puntuación que se lleva. Esta puntuación dependerá del escrito desarrollado, como argumentación por parte del estudiante. La forma cómo utiliza sus conocimientos y cómo los enlaza, lo que construye la solución del problema planteado.

En el caso de las preguntas cerradas se tuvo en cuenta la guía dada por Case y Swanson [21], quienes brindan una serie de estrategias para la construcción de preguntas de opción múltiple. Dentro de las características que consideran más relevantes se tienen:

- El enunciado debe ser claro y no inducir a la posible respuesta.
- Los distractores no deben ser confusos, si bien son menos correctos no deben ser totalmente incorrectos. Esto podría generar en el estudiante la selección de la respuesta verdadera por descarte.
- No usar preguntas que involucren respuestas memorísticas o de simple deducción.
- La pregunta debe evaluar la aplicación de conocimientos, la integración, la síntesis y el juicio acerca de un problema dado.

Considerando estos elementos se construyeron 22 preguntas para la prueba de apropiación y uso de la tecnología y 18 para la prueba de solución de problemas con tecnología. En la siguiente sección se explican los procedimientos utilizados para hacer la validación de las pruebas diagnósticas creadas.

B. Validación de pruebas escritas

La evaluación permite al docente tener una guía de las habilidades que tiene el estudiante, a partir de su aplicación en una situación o entorno. La evaluación de tipo escrito cumple con esta especificación, pero es necesario validarla de tal forma, que proporcione información específica acerca de lo que el alumno conoce y aplica [7]. Por esta razón se analizó si el instrumento evaluativo creado, tiene la capacidad de dar la información adecuada sobre las competencias evaluadas.

Un instrumento efectivo de evaluación debe ser validado de tal forma que haga una medición significativa y adecuada de los rasgos para los cuales fue diseñado [7]. Por eso en este caso se utilizará la validez de contenidos como proceso que permite determinar si las pruebas creadas cumplen con la función para las cuales se diseñaron. Existen diversas formas de hacer esa validez de contenido, puede ser mediante juicio de expertos, quienes revisan si cada una de las preguntas o ítems responden efectiva y coherentemente a las competencias a evaluar [22]. También es posible hacerlo con una aplicación en un grupo piloto y con el análisis de los resultados obtenidos, establecer si el instrumento cumple o no con el

objetivo por el cual se creó [23].

Para validar el contenido de las pruebas diagnósticas diseñadas se realizó la aplicación en un grupo piloto, ya que de esta manera es posible hacer un seguimiento riguroso a partir de los lineamientos tradicionales de la medición y evaluación [23]. Para hacer esta validación se utilizó el índice de dificultad para establecer la proporción de estudiantes que responden correctamente una pregunta de la prueba [24].

$$P_i = \frac{A_i}{N_i} \quad (1)$$

Donde:

P_i es el índice de dificultad de la pregunta.

A_i es el número de aciertos de la pregunta.

N_i es el total de estudiantes participantes.

También se utilizó la validez por índice de discriminación, este indicador es usado para medir habilidades y competencias. Gracias a que brinda información acerca de las posibilidades esperadas en el estudiante que tuvo una puntuación alta en todo el test, partiendo de la suposición que tendrá altas probabilidades de contestar correctamente una pregunta [7]. Lo que permite identificar si la pregunta discrimina o no entre los estudiantes.

$$D_i = \frac{GA_{aciertos} - GB_{aciertos}}{N_{grupomayor}} \quad (2)$$

Donde:

D_i es el índice de discriminación de la pregunta.

$GA_{aciertos}$ es el número de aciertos en la pregunta, por personas con altos puntajes.

$GB_{aciertos}$ es el número de aciertos en la pregunta, por personas de bajas puntuaciones

$N_{grupomayor}$ número total de personas en el grupo más numeroso.

La confiabilidad de una prueba hace referencia a la precisión con que el test mide lo que debe medir, en una población determinada y en las condiciones normales de aplicación [14] y [25]. Esta medida permite corroborar que las pruebas al ser aplicadas en poblaciones similares, producirá resultados consistentes y coherentes con lo que se quiere medir. Para el caso de las pruebas diseñadas se hizo una medición de confiabilidad por consistencia interna, ya que esta brinda la información necesaria a partir de una sola aplicación de las pruebas. Para su análisis se utilizó el Alfa de Cronbach, debido a que analiza la media ponderada de las correlaciones entre las preguntas que forman parte de la escala de medición dada [26].

El último aspecto que se tiene en cuenta para la validación de un instrumento evaluativo es la estandarización. En esta fase se determinan las normas que se considerarán para la aplicación, así como en la calificación e interpretación de los resultados obtenidos. Con estos aspectos será posible determinar que la prueba realizada mide las competencias

planeadas. Lo que permitirá al docente y estudiante identificar las fortalezas y debilidades que se tienen y así desarrollar un proceso educativo mejor planeado [27].

IV. APLICACIÓN DE PRUEBAS Y ANÁLISIS DE RESULTADOS PILOTO

Se aplicaron los instrumentos diagnóstico diseñados con estudiantes de las instituciones I.E.D Julio Garavito Armero de la localidad de Puente Aranda y el Colegio Distrital Clemencia Holguín de Urdaneta de la Localidad Rafael Uribe Uribe de Bogotá. Para la prueba de apropiación y uso de la tecnología se contó con 93 estudiantes y en la prueba de solución de problemas con 50 entre ambas. Los participantes pertenecen al ciclo V de educación secundaria, que corresponde a los niveles de décimo y undécimo.

Las pruebas se aplicaron en el mes de febrero del año 2015, para ello se tomaron dos sesiones de clase de dos horas. En ambas sesiones se aplicó una de las pruebas de forma individual, sin dar a conocer ningún aspecto que condujera a la selección de alguna respuesta. Posterior al proceso de aplicación de las pruebas en las instituciones se hizo el análisis de los resultados obtenidos. En cada una de ellas se analizó validez a partir de los índices de dificultad y discriminación, así como la confiabilidad de cada una de las pruebas.

A. Análisis prueba apropiación y uso de la tecnología

La tabla 1 presenta los resultados de la prueba de apropiación y uso de la tecnología. Para el análisis de los datos se tendrá en cuenta la aceptación de ítems planteada por Guilford [28], donde establece que el rango aceptable para el índice de dificultad es de 0,2 a 0,85. Los resultados son vistos como una curva de distribución de frecuencias, donde las respuestas dadas por el estudiante permiten considerar el nivel de dificultad. Una pregunta que obtiene un 5% puede ser considerada como fácil, con el 20 % medianamente fácil, con el 50 % con dificultad media, con el 20 % medianamente difícil y si tiene el 5 % difícil. En el caso del índice de discriminación el mínimo de aceptación es del 0,3.

Tabla 1. Resultados del índice de dificultad y discriminación de la prueba apropiación y uso de la tecnología

Item	Dificultad	Discriminación	Item	Dificultad	Discriminación
1	0,3	0,45	12	0,64	0,81
2	0,58	0,65	13	0,63	0,84
3	0,61	0,71	14	0,78	0,03
4	0,11	-0,06	15	0,22	0,42
5	0,27	0,36	16	0,61	0,71
6	0,32	0,46	17	0,27	-0,35
7	0,63	0,9	18	0,37	-0,1
8	0,22	0,1	19	0,37	-0,32
9	0,57	0,97	20	0,89	1
10	0,37	-0,16	21	0,02	0,97
11	0,58	0,9	22	0,85	0,45

Las preguntas 4 y 21 tienen índices de dificultad están fuera del rango aceptable definido por Guilford. Puede darse esta situación por ser preguntas demasiado complejas o que podrían no ser claras para el estudiante. Por ello una alternativa podría ser modificarlas o eliminarlas del instrumento. En el caso del índice de discriminación las preguntas 4, 10, 17, 18 y 19, presentan índices de discriminación bajos. Quiere decir que sin importar si pertenecen al grupo de estudiantes de alto o bajo desempeños (GA o GB en la ecuación 2), los escolares pueden responderla de forma correcta o incorrecta. Por tanto la pregunta no discrimina entre los estudiante de alto desempeño y los que tienen bajo nivel. Es por ello que estas preguntas son susceptibles de ser eliminadas, aunque es necesario previamente analizar si afectan o no la confiabilidad de la prueba.

Para el caso de la confiabilidad se analizó el alfa de Cronbach mediante el programa SPSS. También se hizo análisis de escala si se elimina el elemento. Este último muestra qué sucede con la confiabilidad al momento en que se elimina alguna de las preguntas. Los resultados se presentan en la tabla 2 y 3. La tabla 2 muestra que el nivel de confiabilidad es de 0,545, valor que se puede interpretar como una confiabilidad moderada [23]. Esto implica que se tienen preguntas que pueden no contribuir a que la prueba tenga la claridad necesaria para el estudiante. Por tanto habría que analizar que sucede si se eliminan algunos ítems.

Tabla 2. Resultados de la confiabilidad de la prueba apropiación y uso de la tecnología

Alfa de Cronbach	N de elementos
0,545	22

Tabla 3. Resultados de confiabilidad por pregunta de la prueba apropiación y uso de la tecnología

Pregunta	Alfa de Cronbach si se elimina el elemento	Pregunta	Alfa de Cronbach si se elimina el elemento
P1	0,492	P12	0,491
P2	0,486	P13	0,514
P3	0,502	P14	0,472
P4	0,538	P15	0,53
P5	0,547	P16	0,517
P6	0,546	P17	0,606
P7	0,511	P18	0,559
P8	0,509	P19	0,553
P9	0,509	P20	0,51
P10	0,522	P21	0,512
P11	0,522	P22	0,521

La tabla 3 permite establecer que al eliminar algunas preguntas se genera una mejora en el nivel de confiabilidad de la prueba. Por tanto se propone eliminar las preguntas 4, 10, 17, 18 y 19. Es necesario mencionar que al momento que se

eliminan no afecta ningún desempeño. Al hacer la prueba nuevamente, la confiabilidad llega al 0,66 aumentando coherencia y consistencia de la prueba.

De otro lado se concluyó mantener la pregunta 21 en la prueba. Ésta no afecta la confiabilidad de la prueba y su índice de discriminación permite establecer que si diferencia los estudiantes con buen desempeño de aquellos que tienen un nivel bajo. Por tanto se modificó la redacción y se analizó si los conceptos en ella evidenciados evalúan los indicadores esperados. Con esto se espera que la pregunta mejore su validez a partir de su índice de dificultad. Es necesario aclarar que se hará una nueva aplicación del instrumento y con ello volver a medir la fiabilidad, así como los índices de dificultad y discriminación.

B. Análisis prueba solución de problemas con tecnología

La tabla 4 presenta los resultados obtenidos en las 18 preguntas que componen la prueba de solución de problemas con tecnología.

Tabla 4. Resultados del índice de dificultad y discriminación de la prueba de solución de problemas con tecnología

Item	Dificultad	Discriminación	Item	Dificultad	Discriminación
1	0,26	0,35	10	0,21	0,41
2	0,4	0,41	11	0,16	0,24
3	0,34	0,35	12	0,12	0,29
4	0,23	0,41	13	0,52	0,47
5	0,62	0,53	14	0,48	0,29
6	0,62	1	15	0,5	0,47
7	0,34	0,41	16	0,44	0,12
8	0,76	0,41	17	0,52	0,35
9	0,56	0,41	18	0,44	0,3

Nuevamente, a partir de los parámetros de Guilford las preguntas 11 y 12 tienen índices de dificultad fuera del rango. Por tanto es posible establecer que estos ítems son demasiado complejos y deben ser modificados o eliminados de la prueba. En el caso del índice de discriminación, las preguntas 11 y 16, presentan valores fuera del rango. Por tanto estas dos preguntas podrían ser eliminadas o modificadas, por no discriminar entre los estudiantes de buen desempeño y los que tienen bajo nivel. Para decidir si se modifican o se retiran estas preguntas de la prueba se hará un análisis para determinar la manera en que estas preguntas afectan la confiabilidad de la prueba.

Tabla 6. Resultados de la confiabilidad de la prueba de solución de problemas con tecnología

Alfa de Cronbach	N de elementos
0,778	18

Tabla 7. Resultados de confiabilidad por pregunta de la prueba de solución de problemas con tecnología

Pregunta	Alfa de Cronbach si se elimina el elemento	Pregunta	Alfa de Cronbach si se elimina el elemento	Pregunta	Alfa de Cronbach si se elimina el elemento
P_1	0,782	P_7	0,798	P_13	0,768
P_2	0,762	P_8	0,76	P_14	0,76
P_3	0,786	P_9	0,75	P_15	0,744
P_4	0,779	P_10	0,78	P_16	0,749
P_5	0,762	P_11	0,758	P_17	0,76
P_6	0,76	P_12	0,775	P_18	0,72

La tabla 6 presenta un índice de confiabilidad alto para la prueba, lo que quiere decir que el instrumento entregará resultados coherentes y consistentes al momento de ser aplicado. En la tabla 7 al observar el comportamiento de las preguntas 11, 12 y 16 que obtuvieron bajo índice de dificultad y de discriminación, se observa que no afectan de manera sustancial la confiabilidad del instrumento. En el caso de la pregunta 11 se determinó eliminarla, dado que no cumple la función de establecer si el estudiante posee o no la habilidad que evalúa. En el caso de las preguntas 12 y 16 se considera que pueden ser modificadas y con ello mantenerlas en la prueba. Posterior a la modificación de las preguntas, se hará una nueva aplicación y a partir de ella medir los índices junto con la fiabilidad.

V. CONCLUSIONES

Se ha identificado claramente que las dos pruebas validadas deben tener un proceso de corrección en algunos de sus preguntas, para posteriores aplicaciones. Su validación muestra que tienen índices de dificultad y discriminación suficientes en la mayoría de sus preguntas. También al observar la confiabilidad de cada una de las pruebas, evidencian que son lo suficiente coherentes y consistentes para ser aplicadas. Por tanto son instrumentos que al ser utilizados es probable que brinden información coherente y consistente sobre las competencias que evalúan.

Las pruebas diagnósticas validadas son una herramienta evaluativa que brindan a los docentes, información acerca de los conceptos previos que tienen los estudiantes sobre los componentes de “*apropiación y uso de la tecnología*” y “*solución de problemas en tecnología*”. Con esta información el docente podrá hacer una planeación de sus actividades de enseñanza, enfocada en mejorar aquellos desempeños donde se evidencia dificultad. En el caso del estudiante las pruebas le permitirán ver su nivel en cada una de las componentes. Esto contribuye a mejorar sus procesos de aprendizaje, ya que brinda la posibilidad de potenciar aquellos desempeños donde no demuestra mayor nivel de habilidad.

Estos instrumentos y su aplicación adecuada, darán la información necesaria para poder hacer un proceso de planeación de clases más coherente con las necesidades del grupo. Lo que puede llegar a generar mejores índices de

desempeño y mayor motivación por parte del estudiante, frente a los conceptos que se presentan.

Como trabajo futuro se proyecta implementar los instrumentos contruidos y validados en varias instituciones educativas en Colombia, de modo que se conozca su pertinencia en las aulas de clase. Durante este proceso se realizará una evaluación de uso por parte de los docentes y estudiantes, generando un conjunto de recomendaciones y sugerencias frente a su diseño y aplicación. A continuación se implementarán las sugerencias y recomendaciones que surjan del proceso anterior y se desarrollará una nueva versión de dichos instrumentos. Por otra parte se proyecta desarrollar una herramienta de software que facilite su uso por parte de los docentes además de simplificar el procesamiento e interpretación de la información resultante de la implementación de los instrumentos en clase.

RECONOCIMIENTO

Este trabajo se desarrolla en el marco del programa de investigación (Ctr No. 0826-2013) que se financia con recursos del Patrimonio Autónomo Fondo Nacional de Financiamiento para la Ciencia, la Tecnología y la Innovación, Francisco José de Caldas, aportados por el Ministerio de Educación Nacional.

REFERENCIAS

- [1] Bakx, A, Baartman L & Schilt-Mol (2014). Development and evaluation of a summative assessment program for senior teacher competence. *Studies in Educational Evaluation* 40, 2014, 50 -62.
- [2] Tobón, S. (2008). La formación basada en competencias en la educación superior: el enfoque complejo. *México: Universidad Autónoma de Guadalajara*.
- [3] Kimbell, R (2012). Understanding assesment its important; its danger; its potential. Sense Publishers, 137 – 169
- [4] Diaz, F (2006). Enseñanza situada: Vínculo entrela escuela y la vida. Mc Graw Hill
- [5] Kopaiboon, W, Reungtrakul, A & Wongwanich, S (2013). Developing the Quality of ICT Competency Instrument for Lower Secondary School Students. 5th World Conference on Educational Sciences - WCES 2013
- [6] Ritzhaupt, A & Florence M (2013). Development and validation of the educational technologist multimedia competency survey. Association for Educational Communications and Technology 2013
- [7] Froncek, B, Hirschfeld, G & Thielsch, M (2014). Characteristics of effective exams Development and validation of an instrument for evaluating written exams. Elsevier 513 – 522.
- [8] Arsenault, J, Godsoe, S, Holden, C, & Vetelino, J. (2005) Integration of sensors into secondary school classrooms. pages S2F –33, oct. 2005
- [9] Kocijancic, S (2001). Mechatronics as a challenge for teaching technology in secondary education. 1:T2E –1–4 vol.1, 2001.
- [10] Excoffon, E, Papillon, F, Fesquet, L, Bsiesy, A & Bonnaud, O (2012). New pedagogical experiment leading to awareness in nanosciences and nanotechnologies for

- young generations at secondary school. pages 1–4, June.
- [11] William, P. Technology education for teachers. International technology education studies, Sense Publishers (2012)
- [12] Min. Educación. (2008) Serie 30: Ser competente en tecnología: Una necesidad para el desarrollo. Serie 30, 1, 2008.
- [13] Martínez, M., Amante, B., Cadenato, A., & Gallego, I. (2012). Assessment tasks: center of the learning process. *Procedia-Social and Behavioral Sciences*, 46, 624-628.
- [14] Almanera, J & Llorente M. The experts judgment application as technic evaluate the information and communication technology. *Revista de Tecnología de Información y Comunicación en Educación*. Volumen 7, N° 2 Julio-Diciembre 2013.
- [15] Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into practice*, 41(4), 212-218.
- [16] Martínez, M., Amante, B., Cadenato, A., & Gallego, I. (2012). Assessment tasks: center of the learning process. *Procedia-Social and Behavioral Sciences*, 46, 624-628.
- [17] Gatica-Lara, F., & Uribarren-Berrueta, T. D. N. J. ¿ Cómo elaborar una rúbrica?. September, 15, 2014: http://riem.facmed.unam.mx/sites/all/archivos/V2Num01/10_PEM_GATICA.PDF
- [18] OCDE (2006). PISA 2006: Marco de la evaluación. Conocimientos y habilidades en ciencias, matemáticas y lectura. PISA 2006. Disponible en: <http://www.oecd.org/pisa/39732471.pdf>
- [19] Stangal, B & Kstner (2011). Multiple Choice and Constructed Response Tests: Do Test Format and Scoring Matter? *Procedia-Social and Behavioral Sciences*, 12 263 – 273.
- [20] Husain, H, Bais, B, Hussain, A & Samad S (2012). How to construct open ended questions. *Procedia-Social and Behavioral Sciences*, 456 – 462.
- [21] Case, S & Swanson, D (2006). Cómo elaborar preguntas para evaluaciones escritas en el área de ciencias básicas y clínicas. Tercera edición, National Board of Medical Examiners 2006
- [22] Corral, Y (2009). Validez y confiabilidad de los instrumentos de investigación para la recolección de datos. *Revista de las ciencias de la educación*. Vol 19, N 33, enero – junio.
- [23] Obando, P (2009). Construcción y validación de una prueba para medir conocimientos matemáticos. *Horiz. Pedagógico*. Volumen 11. N° 1. 29- 37.
- [24] Backhoff, E, Larrazolo, N & Rosas M (2000). The Level of Difficulty and Discrimination Power of the Basic Knowledge and Skills Examination (EXHCOBA). *Revista Electrónica de Investigación Educativa* Vol. 2, No. 1
- [25] Aiken, L. (1996). *Tests psicológicos de evaluación*. México: Prentice-Hall.
- [26] Merino C & Lautenschlager, G (2003). Comparación Estadística de la Confiabilidad Alfa de Cronbach: Aplicaciones en la Medición Educativa y Psicológica. *Revista de Psicología de la Universidad de Chile*. Vol. XII, N° 2: Pág. 127-136.
- [27] Saxton, E, Burns, R, Holveck S, Kelley, S & Prince, D (2014). A common measurement system for k-12 STEM education: adopting an educational evaluation methodology that elevates theoretical foundations and systems thinking. Elsevier. *Studies in educational evaluation* 40, 18 -35.
- [28] Guilford, J. P. (1955). Varieties of creative giftedness, their measurement and development. *Gifted Child Quarterly*, 19, 107–121.