

IMPLICACIONES EDUCATIVAS DE LA MEDICIÓN DEL APRENDIZAJE DE LA FÍSICA EN LAS UNIVERSIDADES: UN ANÁLISIS BAJO LA PERSPECTIVA DE LA ESCRITURA COMO PROCESO COGNITIVO¹

Edilberto Arteaga-Narváez, MS. Ed.D.

Catedrático Asociado
Ciencias y Tecnología
Recinto Metropolitano
Universidad Interamericana de Puerto Rico

Will Alfredo Gómez Uribe, MS. Ph.D.

Catedrático Auxiliar
Departamento de Ciencias Física
Recinto San Juan
Dewey University

Resumen

La evaluación de los cursos está determinada por criterios de medición sujetos a ciertas pruebas: de pregunta abierta, de solución de problemas o de pruebas de selección múltiple, principalmente. Se pasa por alto procesos de meta cognición en la evaluación, sin embargo la evaluación del aprendizaje implica cierto grado de reflexión acerca de cómo aprenden y en muchos casos un análisis de qué aprenden (Lucero & Meza, 2009). Este proceso de evaluación se da a través del tiempo, lo que permite instancias permanentes de reflexión.

En este estudio se realiza un análisis de ítems, tradicionalmente y bajo la perspectiva de la escritura (Carlino, P. 2005) y el lenguaje de los ítems para los exámenes de Física General suministrados en una universidad en Puerto Rico.

¹ Ponencia presentada en el II Congreso de Investigación, Educación y Formación de Docentes. Medellín, Colombia, 26 de marzo de 2009.

En esta ponencia se resalta, a modo de conclusión, aspectos como: (1) la utilización de las pruebas de selección múltiple en los exámenes con una frecuencia notoria en la medición del conocimiento, (2) el análisis de ítems refleja que la prueba presenta un grado pobre de validez, (3) los análisis cualitativo de los ítems revelan información de los procesos cognitivos del binomio profesor-estudiante y del lenguaje utilizado en la redacción de los ítems de la prueba.

Palabras Claves: Física, análisis de ítems, lenguaje.

Abstract

The evaluation of the courses is determined by measurement criteria related to certain tests: open-ended questions, solution of problems or tests of multiple choice questions. Metacognition in the evaluation is ignored goal processes, nevertheless participants' evaluation of their learning implies a certain degree of reflection about how they learn and in many cases an analysis of what they learn (Lucero & Meza, 2009). This process of evaluation occurs over time, which allows "permanent" instances of reflection.

In this study a traditional analysis of items is made, and under the perspective of the writing (Carlino, P. 2005) and the language of the items for the provided examinations of General Physics in a university in Puerto Rico.

In this speech it is emphasized, as a conclusion, aspects like: (1) the use of multiple choice questions in student examinations with a well-known frequency in the measurement of the knowledge, (2) the analysis of items reflects that the test shows a poor degree of validity, (3) the qualitative analyses of the items reveal information about the cognitive processes of the binomial teacher-student and the language used in the test items.

Key words: Physics, item analysis, and language.

Introducción

Antecedentes

Uno de principales retos para la tarea docente es medir el aprovechamiento académico de los estudiantes. Las pruebas de aprovechamiento son una importante herramienta para medir el conocimiento y las destrezas que los estudiantes poseen en un particular área de contenido (Zhang, 2008). La pregunta necesaria es ¿en qué medida los profesores poseen preparación acerca de cómo hacer esta medición? La pregunta es válida si se tiene en cuenta que la mayoría de las pruebas educacionales adoptan ítems de selección de respuestas y de construcción de respuestas. Con estas pruebas en el pasado se cubrían grandes cantidades de contenido y más recientemente para valorar la profundidad en las destrezas de pensamiento, en adición a esto se incorporan una gran variedad de estímulo en el ítems, como gráficas, textos, tablas, diagramas, dibujos y mapas (Zhang, 2008), que permiten un mayor espectro de información en lo que se quiere comunicar.

Trasfondo Problema

Uno de los asuntos que puede señalarse de esta práctica es el hecho de que ¿con qué frecuencia estos instrumentos están midiendo lo que se desea medir? Así como menciona Zhang (2008) que las respuestas a los ítems puede ser afectada no sólo por el blanco cognitivo para el que fue construido, sino también por dimensiones latentes secundarias tales como: la habilidad para procesar varios estímulos o producir diferentes tipo de resultados. Una cuarta característica que creemos guarda relación con lo anterior es el lenguaje en que están escritos los ítems, tanto como un mecanismo de comunicación como un mecanismo cognitivo. Al referirnos al lenguaje, no nos referimos, necesariamente, a la lengua o idioma, sino al conjunto de símbolos y códigos propios de las ciencias en particular de la física. Rincke (2006) señala que existe un lenguaje especial cuando los estudiantes aprenden física, más allá del aprendizaje de los

conceptos o ideas nuevas que debe aprender en un curso de física. De manera que si este lenguaje coexiste con el aprendizaje de conceptos de física, debe atenderse también en los procesos evaluativos, específicamente en la medición del aprendizaje mediante exámenes de selección múltiple, los cuales son ampliamente utilizado por los profesores universitarios que enseñan cursos de física general. La forma como están escritos los ítems en estos exámenes, las expresiones lingüísticas y simbólicas utilizadas, son factores inherentes al proceso de construcción de la prueba, por lo que deben tenerse en cuenta al momento de entender la naturaleza de estos exámenes.

La evaluación de los cursos, queda estrictamente determinada por criterios de medición que están sujeto a ciertas pruebas, que en su mayoría son pruebas de pregunta abierta, de solución de problemas como aplicación de los principios o leyes estudiadas o de pruebas de selección múltiple, pobremente construidas en muchos casos. Generalmente estos aspectos forman parte de lo que algunos llaman el prontuario o silabo del curso. Este prontuario llega a transformarse prácticamente en un contrato entre los estudiantes y la institución que imparte educación, a través del profesor a cargo de la clase. Dado que la evaluación del aprendizaje implica cierto grado de reflexión acerca de cómo aprenden y en muchos casos un análisis de qué aprenden (Lucero & Meza, 2009), el mismo debe darse a través del tiempo lo que permite instancias permanentes de reflexión. De manera que al momento de construir los exámenes se tenga una visión más holística de los aspectos relacionados a la construcción de la prueba, entre ellos la escritura o lenguaje especial de la materia a evaluar.

Preguntas de investigación

La problemática anterior permite plantearnos la siguiente pregunta de investigación:

¿En qué medida el lenguaje y la forma en la construcción de una prueba de aprovechamiento, tipo selección múltiple, es tenido en cuenta en la medición del aprendizaje de los estudiantes en un curso de Física de nivel universitario?

Esta interrogante da pie a las siguientes Sub-preguntas que sirvieron de guía en este estudio:

¿Hasta qué punto es válida esta prueba de selección múltiple utilizada para medir el aprendizaje de los estudiantes de un curso de Física de nivel universitario?

¿Cuáles son las características en el lenguaje utilizado para construir la prueba que contribuyen a la confiabilidad, el índice de discriminación y el índice de dificultad?

¿En qué medida la forma cómo están escritos los ítems en la prueba atienden el lenguaje especial de la física?

Objetivos

Entre los objetivos a alcanzar en este estudio tenemos:

- 1) Dar un primer paso en la documentación de las formas de evaluar los cursos de Física General de nivel universitario; utilizando pruebas tipo selección múltiples y desde una perspectiva del lenguaje especial de la materia.
- 2) Identificar la relación entre el lenguaje utilizado en las ciencias, específicamente en la Física, y la construcción de instrumentos de evaluación.

Fundamentación Teórica

Existen algunas maneras de medir el aprendizaje en las ciencias, entre las más usadas están: preguntas abiertas, resolución de problemas y pruebas de selección múltiples. Sin embargo, las preguntas abiertas poseen un elemento de

subjetividad de parte del corrector, lo que da lugar a una multiplicidad de problemas que restan calidad a la evaluación (Aliss, 2003) Asimismo, la resolución de problemas en ciencias está dirigida exclusivamente a la solución de ejercicios prediseñados, con cuya solución el profesor infiere cuanto han aprendido los estudiantes. Esta característica presente en la solución de problemas de esta manera, no se ajusta a la concepción de solución de problemas establecidos en la literatura del tema. Por otro lado, la “objetividad” en las pruebas objetivas tiene connotaciones equivocadas (Morales, 2007) ya que si bien la corrección es objetiva (una respuesta o está bien o está mal), la formulación de la pregunta (qué y cómo se pregunta) como en el criterio donde se coloca el punto mínimo por encima del cual caen los estudiante aptos o capaces son decisiones subjetivas de los profesores. De manera que, partiendo de esta aclaración puede ser fructífera una mirada a la práctica didáctica de la medición del aprendizaje en ciencias, específicamente en curso de física, bajo la perspectiva del lenguaje especial de la misma física.

La escritura como medio necesario en la construcción de ítems

La escritura es un elemento estrictamente relacionado con el lenguaje. Desde la perspectiva sociocultural la educación es considerada una actividad social y socializadora mediatizada por signos (Iglesias y De Micheli, 2009), así lo señalan los escritos de Vigotsky. De manera que dentro de la construcción de ítems, hacemos uso de ciertos signos y ciertos términos que nos permiten comunicar información. Es decir, la escritura es lenguaje en esencia dirigido a una persona que no está o es imaginaria o a nadie en particular (Bouzas, 2004), como el caso cuando el profesor escribe una prueba para medir el aprendizaje de sus estudiantes. Esta escritura no se hace en pro de un estudiante en particular, la prueba es para cualquier estudiante que esté tomando el curso o que tomará el curso. El ejercicio de la escritura requiere de la utilización de términos y sus significados, tanto del idioma en que se escribe como del campo o materia en que se escribe. Por ejemplo, términos como medición que consiste en reglas para

asignar números a los objetos, con el propósito de representar cantidades de atributos (Nunnally, 1987), o el concepto de trabajo, que puede traer implicaciones de significados distintas dentro del idioma como lenguaje, como dentro del lenguaje especial de la física.

Es importante resaltar aquí que cuando medimos el aprendizaje, estamos midiendo un atributo psicológico, para los cuales las reglas o procedimientos de medición no son tan obvios. Ahora bien, si el profesor no tiene en cuenta este detalle podría hacer uso indiscriminado de la terminología de las ciencias y “asumir” o dar por sentado que los estudiantes no sólo conocen el contenido, sino también los términos, códigos y símbolos propios de la estructura lingüística de las ciencias o la física. De modo que, al medir el aprendizaje no debe perderse de vista que medir implica un proceso de abstracción (Nunnally, 1987) y esta abstracción se hace sobre el atributo que se desea medir. Así que la coherencia entre objetivos del curso y los procedimientos de medición tienen que estar alineado. En este punto se hace indispensable el uso de alguna taxonomía de objetivos cognitivos (como la de Bloom) que de dirección a la abstracción que pretendemos con la medida del aprendizaje en unas de sus dimensiones (ejemplo comprensión), con determinado instrumento. En resumen, la medición siempre concierne a un atributo en particular, que nos obliga a una cuidadosa consideración de la naturaleza del atributo antes de intentar su medición (Nunnally, 1987).

Uno de los elementos determinantes de la naturaleza del atributo (aprendizaje de conceptos de física en nuestro caso), es la misma composición del lenguaje propio de la física. Por ejemplo, términos como “*la fuerza actúa sobre...*”, puede llevar a interpretaciones del concepto desde distintos puntos de vista y no necesariamente desde el punto de vista que el profesor quiere. Ya que, sólo los expertos en la materia (digamos profesores de física) tendrán la profundidad de análisis por encima de los novatos (estudiante). Este fenómeno es reportado en la literatura por Iglesias y De Micheli (2009), cuando dicen que por ejemplo la

expresión: *“los cactus se han adaptado a los ambientes con poco agua”*, se refiere a un modelo científico diferente a *“los cactus están adaptados a los ambientes con poco agua”*. Recalcan que sólo la segunda expresión se refiere al modelo de evolución por selección natural validado actualmente.

Metodología

Diseño

Este es un estudio de caso con técnicas mixtas de recolección de información. Se consideró como caso las implicaciones que tiene una prueba de aprovechamiento, tipo selección múltiple, sobre la medición del aprendizaje.

Participantes

En la primera fase de este estudio se utilizaron 4 secciones del curso para un total de 104 estudiantes. Se seleccionó una serie de expertos: 2 expertos en construcción de pruebas, un experto en contenido, y un experto en redacción. Para la segunda fase se procedió a corregir este examen a tendiendo las recomendaciones dada por los expertos en la primera fase y se seleccionaron 3 profesores de física con una experiencia de por lo menos 5 años ofreciendo este tipo de cursos.

Técnicas de recopilación de Información

En este estudio se utilizó como técnicas de recopilación de información, el análisis de ítems a un examen tipo selección múltiple, usando la teoría clásica de ítems y la entrevista en profundidad (Fraenkel & Wallen, 2000; Lucca & Berríos, 2003; Punch, 1999; Seidman, 1998).

Procedimiento

Se seleccionó una universidad en Puerto Rico, y se tomaron dos semestres académicos, durante el cual se estuvo recopilando información, sometiendo el examen a los estudiantes y entrevistando al grupo de profesores participantes.

Análisis de Información

Se hizo un análisis de ítems con el programa SPSS para la parte cuantitativa y como método de análisis de la información recogida en las entrevistas semi-estructuradas se utilizó el postulado por Wolcott (1994), conocido como DAI (Descripción, Análisis e Interpretación). La mayoría de las categorías de análisis emergieron del análisis de la información. Se organizó la información identificando patrones para crear categorías.

Resultados, análisis y discusión

Para fines de claridad en el análisis y discusión de los resultados y dado que presentamos dos fuentes de información de naturaleza distinta. Hemos decidido hacer el análisis guiado por las sub-preguntas de investigación.

¿Hasta qué punto es válida esta prueba de selección múltiple utilizada para medir el aprendizaje de los estudiantes de un curso de Física de nivel universitario?

La media aritmética del examen fue de 17, que aproximadamente es la mitad del total de ítems (29 puntos). La desviación estándar (σ), hallada para este examen fue de 4.98 (Ver tabla 1), esto nos indica, la dispersión de las puntuaciones de los estudiantes de la media aritmética. De manera que, mientras más grande sea este número, mayor heterogeneidad existe en la medición del aprendizaje que se espera que el grupo tenga.

Tabla 1: Resultados generales de la prueba de aprovechamiento.

Ítems	Ítems analizados	Muestra	σ	α	Media	Mediana
30	29	104	4.98	0.79	17	17.6

En la gráfica 1 se observa que la mayor frecuencia en la puntuación de los estudiantes, 24 de ellos obtuvieron puntuaciones entre 50% y 60%, es decir aproximadamente 23% (24/104). Esto corresponde a notas prácticamente de F

respecto a la escala utilizada ($X < 60$). Ahora, el porcentaje de estudiantes que lograron pasar la el examen con C ($X > 70$) o más fue de 29% (30/104). Si consideramos sólo esta información estadística especularíamos acerca de lo difícil que pudo haber sido el examen, como una explicación al fracaso de este grupo de estudiantes. En este punto no se tiene claro cuáles fueron las razones a las que se debieron estos resultados. Veamos un parámetro más para el análisis: el índice de dificultad.

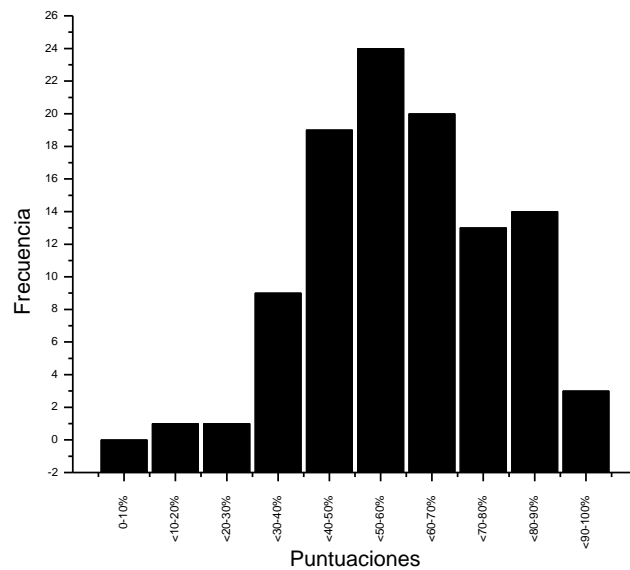


Figura 1. Distribución de frecuencia de puntuaciones

El índice de dificultad de toda la prueba es la razón entre la media aritmética y el número total de ítems. Para este examen fue de 0.57 o 57%. De manera que la prueba se encuentra en términos promedios de dificultad, es decir un 50% (Vera, 2004).

Si consideramos únicamente este parámetro estadístico, podríamos clasificar la prueba como aceptable respecto al grado de dificultad. No obstante, los resultados en las puntuaciones de los estudiantes mostrados en la figura 1, son contradictorios a estos resultados. Miremos en mayor detalle, la dificultad de la prueba. En la figura 2 se muestra un la relación entre el índice de dificultad

respecto a los ítems. El índice de dificultad del ítem es la razón entre la cantidad de estudiantes que respondieron correctamente y el número total de estudiantes que contestaron el examen. Si esta razón se aproxima a 1.00, indica que el ítem es fácil y en caso contrario que se acerque a 0, el ítem es muy difícil. Observamos que 17 de los 30 ítems están por encima del 0.6 lo que indica que estos ítems son de mediana dificultad. Con más detalle se observa que la mayoría de los 17 ítems de mediana dificultad están alrededor de 0.8, es decir entre los fáciles, 59% (10/17) (Verdejo y Medina, 2008).

Lo anterior muestra que la prueba fue relativamente fácil, sin embargo los resultados obtenidos por los estudiantes no fueron los mejores, puesto que 72% de los estudiantes fracasaron en la prueba (sólo 30 estudiantes obtuvieron calificaciones igual o mayores de 70 puntos en 100). Una posible explicación recae en la forma lingüística en que estaban escritos los ítems. Esto es cónsono con lo señalado por Rincke (2006), respecto a la presencia de un lenguaje especial en el aprendizaje de la física. De igual manera, es pertinente tener en consideración la correlación de ítem o discriminación del ítem (una medida de diferenciación entre los que saben más y los que saben menos).

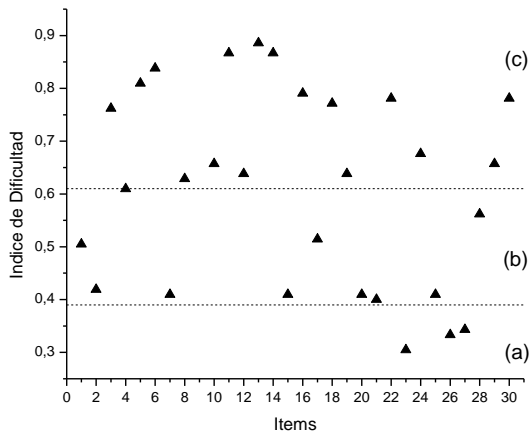


Figura 2. Índice de dificultad vs. Ítems

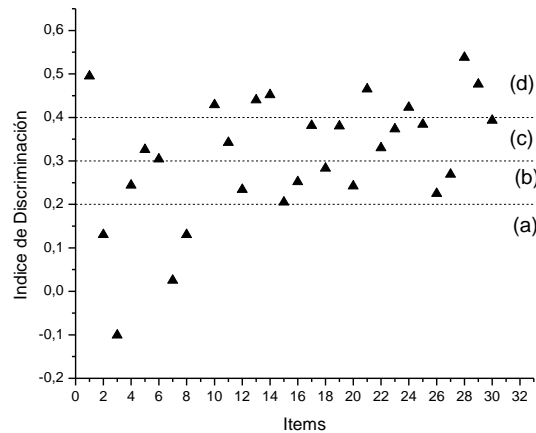


Figura 3. Índice de discriminación vs. Ítems

En la figura 3, observamos que la mayoría de los ítems presentan índice de

discriminación menor a 0.4 (Ver figura 3). Es decir que los ítems discriminan, pero no son los mejores. Esto se puede explicar dado que, no necesariamente, si un ítem presenta un índice de discriminación bajo, esto no es un indicativo de la calidad del ítem. Puede ocurrir que a índices de discriminación bajo correspondan a índices de dificultad tanto altos como bajos. Así que cuando la pregunta es muy fácil o muy difícil el índice de discriminación puede ser bajo y en estos casos se podrían tomar otras decisiones.

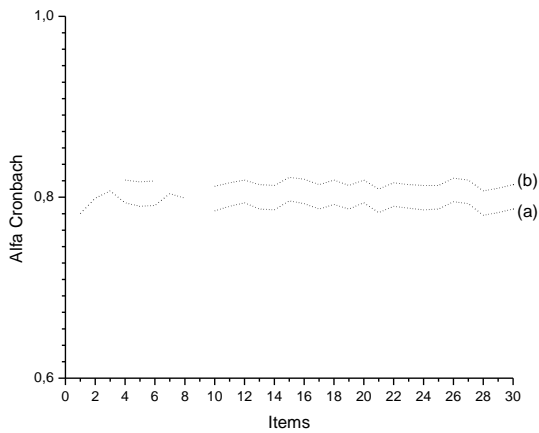


Figura 4. Índice de confiabilidad vs. Ítems. (a) Todos los ítems incluidos, (b) Algunos ítems excluidos

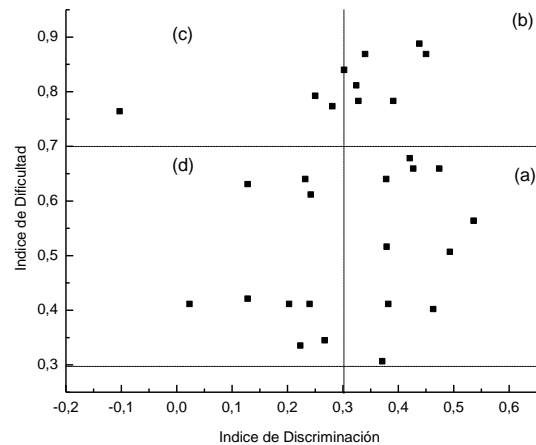


Figura 5. Índice de dificultad vs. Índice de discriminación.

La figura 4, representa la confiabilidad de la prueba excluyendo el ítem correspondiente. En la figura 4(a) se grafica la confiabilidad para 29 ítems, y en la figura 4(b) se grafica la confiabilidad para 25 ítems (ítems con índice de discriminación menores de 0.25 no fueron tenidos en cuenta). Este valor de alfa de Cronbach (α) fue de 0.79 (Ver tabla1) para la parte (a) y de 0.82 para la parte (b). Note que la diferencia entre estos valores no es relevante (0.3). A pesar que estos números se encuentra justo por debajo de los parámetros considerados altos, podríamos decir que éste es aceptable, ya que la literatura reporta como valores aceptables, los comprendido entre 0.80 y 0,90 para pruebas de selección múltiple

(Nunnally, 1987; Adkins, 1981; Vera, 2004; Verdejo y Medina, 2008).

En la figura 5, los 30 ítems son graficados respecto a los índices de dificultad y el índice de discriminación. El área demarcada con la letra (a), representa los ítems con índices de discriminación mayores de 0.30 y nivel dificultad entre 0.30 y 0.70. Estos ítems serían los mejores candidatos a tener en cuenta para una futura prueba de aprovechamiento correspondiente este tema. Pero son sólo 7 (tomando los más cercanos al centro del rectángulo), lo cual representa un 23% de ítems verdaderamente bueno.

En síntesis, la prueba presenta un pobre proceso de validez, por las siguientes razones, sólo el 23% de los ítems mencionados en el párrafo anterior son los que verdaderamente se pueden rescatar para ser considerados en una prueba que se perfila como válida como instrumento de medición del aprendizaje de los conceptos evaluados en física para este periodo.

¿Cuáles son las características en el lenguaje utilizado para construir la prueba que contribuyen a la confiabilidad, el índice de discriminación y el índice de dificultad?

Para obtener información respecto a esta pregunta, hicimos uso principalmente de la entrevista semi-estructurada, realizada a tres profesores con experiencia en la enseñanza de cursos de física similares al que corresponde este instrumento. Así que, entre las categorías que surgieron del análisis de estas entrevistas, veamos en dos de ellas: “*lenguaje ambiguo*”, “*formación conceptual previa en física*” y dos sub-categoría que pertenecen a la primera categoría mencionada anteriormente, estas son “*información incompleta*” y “*lenguaje de la física*”,

La categoría que con más frecuencia se destacó fue la del lenguaje ambiguo. Esta categoría se refiere a cualquier ambigüedad señalada por los participantes en los ítems, en término de la comprensión del mismo. Varios ítems

presentaron esta característica, por ejemplo uno de los profesores, manifiesta al respecto que:

“...habían preguntas que no..no estaban bien claras, cuestiones de conceptos, eh, descripción del problema, fue lo más...no era clara la pregunta de lo que estaban pidiendo [en el examen]”

Debemos señalar que la ambigüedad en la interpretación del examen puede deberse a varios factores, no obstante, estamos focalizado en los aspectos relacionado con la forma cómo esos ítems están escritos. De manera que entre esas razones para encontrar el examen confuso, tocaremos la ambigüedad en el lenguaje, específicamente en los términos, palabras y hasta conceptos utilizados, uno de los profesores expresaba lo siguiente al respecto:

“...hay algunas palabra que producen.. o me produjeron confusión o me llamaron la atención”

“...cuando usa la palabra traslación molecular, cuando el calor pasa, eh, traslación está en el ítem 1 pasa en el ítem 2”

E: Cuando escucha traslación que viene a su mente

P: movimiento...

E: ¿y no es a eso que se refieren?

P: si, pero [silencio] yo lo puedo entender, pero le reconozco cierto grado de dificultad para entender la palabra...alguna persona que no haya tomado, eh, que no tenga un conocimiento previo, específicamente cuando dice: la energía cinética de la traslación molecular, como que...se, se puede quedar un poco en blanco la persona...¿qué en realidad me quiso decir?”

Lo anterior reafirma, lo señalado por Rincke (2006), acerca de la existencia de un lenguaje especial de la física. Es decir, que más allá de la semántica del idioma subyace elementos propios de la física, que deben ser asimilados previa y

hasta paralelamente a los conceptos e ideas nuevas aprendidas. Otro profesor señaló que no es que los términos te lleven a interpretaciones diferentes, sino que te pone, por decirlo así, difícil el proceso de entender lo que se pregunta:

“no, no es una interpretación diferente, sino que le dificultaría, el, el proceso de entender la premisa, no es que la vayan a mal interpretar”

De igual manera términos o palabras como “pasa” como pueden presentar cierto grado de ambigüedad, debido a lo poco formal que es para el lenguaje de la física:

“E: ¿qué pasa con el término ‘pasa’?”

P: si lo que pasa es que transfiere [silencio], lo que pasa es que transfiere, pasa o pasar, eh, el calor pasa...de..el calor pasa de un objeto a otro, no sé, fue que me creo más incomodidad que duda yo hubiese usado mejor transferir...”

La palabra tolera cae bajo el mismo conflicto encontrado con el término anterior:

P: en el tres [ítems] dice tolere, yo encontré dificultad en término de la redacción.

E: me puede explicar eso.

P: si, eh, se hace referencia a un dibujo y a unas viga de 15 metros de largo, de un puente para que tolere de una temperatura, eh, yo explicaría mejor en términos de la situación y que hay unas fluctuaciones en temperatura, unos cambios de temperatura y se quiere de que no colapse, de que no se precipite, que es lo mismo que es lo mismo que tolerar [silencio], pero no es tan específica la palabra tolere para ver ese rango, o el rango de acción en esas temperaturas...”

Existes, de igual manera conceptos que al utilizarse inadecuadamente pueden traer confusión para quien lo lee:

“...transporte de material caliente, no sé por qué adjudicar el, el adjetivo caliente que me parece redundante y se presta a confusión ya que estas hablando [en el examen] de transferencia de calor... que se presume de que hay un proceso térmico, por lo menos yo lo asumo...”

El lenguaje especial de la física para rescatar el término de Rincke (2006), establece ciertos pautos o estilo, al cual debe seguir todo profesor que enseñe la materia. De modo que, salirse de este estilo presupone una complicación para interpretar lo que leemos en un examen de física, sin tener conocimientos previos en este campo. Por ejemplo el uso de prefijo posee ciertas características que hay que tener en cuenta. Uno de los profesores llamo la atención en el siguiente sentido:

“...en el ejercicio 19 ponen la palabra mega es igual a uno por diez a la seis, ¡fantástico!, pero nunca se pone la palabra mega, se utiliza la eme mayúscula...si en realidad quisiera darle esa ayuda al estudiante debiera colocar la palabra mega para saber que significa”

En cuanto a la preparación o conocimiento previo en el campo, como medio que garantice la adecuada interpretación de los exámenes cuando leemos su contenido, este estudio encontró la categoría “*formación conceptual previa en física*”, la cual consiste en cualquier cuerpo de información en la física, matemática o cualquier otro campo afín, que previamente haya sido adquirido y que contribuye a la interpretación del examen. En este estudio se encontró que los profesores entrevistados, reconocen este hecho. Uno de los profesores manifestó al respecto:

“...no si yo no hubiese tomado un curso de física básico general, hay algunas preguntas que no...[contestaría] se le haría difícil a alguien que no

tenga formación en física contestarla porque hay algunas palabras claves que se discuten...[enseñan]”

Otro profesor dijo que los estudiantes que estén en adelantados en años escolares tendrían cierto beneficio frente a este examen de física, cuando se le preguntó si creía que los estudiantes que tomaran este examen sacarían resultados similares al que él obtuvo:

“...posiblemente...dependiendo del año de preparación académica en que este...si, si dependiendo si esta en cuarto año, depende de la formación que tenga previa...”

Otro, manifestó lo mismo, pero de una manera un tanto incrédula, respecto a la certeza de que los estudiantes tengan los conocimientos suficientes para contestarlo como él:

“...yo no creo que un estudiante de ingeniería te resuelva [el examen] de la misma manera que yo [obtener el mismo resultado], yo conozco los temas me entiendes...yo conozco los temas.”

En síntesis, creemos que los resultados obtenidos en la confiabilidad, el índice de discriminación y el índice de dificultad, pudieron deberse en gran medida a la forma cómo estaban escritos los ítems. Es decir, poco se tuvo en cuenta el lenguaje especial de la física y quizás el marco de referencia tomado por quienes hicieron el examen fue un criterio oculto que permitió el énfasis o no de la información, en cierto ítems.

¿En qué medida la forma cómo están escritos los ítems en la prueba atienden el lenguaje especial de la física?

Los resultados expuestos hasta el momento, pueden llevarnos a consideraciones donde el fracaso de los estudiantes se deba a la construcción de la prueba en términos de validez, confiabilidad y discriminación. Sin embargo, un punto crucial a considerar es la escritura bajo la cual la prueba fue elaborada.

Elementos como la “expresión lingüística”, las expresiones matemáticas” y “expresiones simbólicas” que fueron categorías encontradas en las entrevistas con los profesores participantes, pudieron tener implicaciones didácticas en este proceso de medición. El manejo de símbolos en la física como en la matemática presenta dos vertientes en el lenguaje utilizado. Esto es, el simbolismo matemático manifestado a través de la formulas y el simbolismo relacionado con los conceptos físico estudiado. Estos últimos forman parte del lenguaje especial de la física que señala Rincke (2006), que los estudiantes deben aprender además de las ideas y conceptos nuevos. De manera, que es posible encontrar a un estudiante que perfectamente resuelve un problema de física en su forma matemática, sin saber nada profundo del concepto (Del Amo, 1989). Eso abre camino para reflexionar en torno a los señalado por Rodríguez (2003) que un proceso de evaluación debe ser practicado con el fin de utilizar sus resultados para ayudar a establecer metas adecuadas y a descubrir métodos más efectivos destinados a lograr esas metas, y no para recompensar o castigar. Nuestra experiencia en la enseñanza de la física en las universidades y respecto a esta forma de medir el aprendizaje es que el aprendizaje se infiere de los resultados obtenidos de los ítems que se contestaron correctamente. En este sentido se procede similar a lo expuesto por Rodríguez (2003), por recompensa y castigo, cuando quizás los resultados en la prueba no tengan mucho que ver con el aprendizaje, debido a la pobreza en los mecanismos de control de calidad del instrumento.

Conclusiones e Implicaciones

Dado que confiabilidad de 0.79, el índice de dificultad promedio de la prueba fue de 0.57, que son valores que tomados individualmente, no arrojan información competente respecto a la calidad del instrumento de evaluación.

Una de las posibles razones fue el desconocimiento del lenguaje especial de la física, como elemento fundamental en el proceso de enseñanza y aprendizaje de la física. Este desconocimiento del lenguaje interno de la física,

pudo contribuir a la ambigüedad encontrada por los profesores esta prueba para la medición del aprendizaje. Si se desconocen tanto el lenguaje interno de la física, como la preparación previa que trae el estudiante, podríamos caer en una medición “injusta” y desatinada del aprendizaje del estudiante.

Otra posible razón fue la no existencia de un proceso de validez para el instrumento. Aunque no todos los ítems resultaron ser atractivo para seguir siendo utilizados en futuros exámenes, vale la pena tenerlos en cuenta, ya que una revisión tanto de las preguntas como de las alternativas utilizadas podría ir mejorando la calidad del examen.

Referencias

- Adkins, D. (1981). *Elaboracion de tests: Desarrollo e interpretación de aprovechamiento*. Editorial Trillas, Mexico.
- Aliss, E. (2003). Búsqueda de la calidad de los procesos de medición de los aprendizajes. *Acta Nova*. Vol. 2, (3), diciembre 2003 pp. 409-417.
- Bouzas, P. (2004). *El constructivismo de Vigotsky: Pedagogía y aprendizaje como fenómeno social*. Longseller. Buenos Aires, Argentina.
- Carlino, E. (2005). *Escribir, leer y aprender en la universidad: una introducción a la alfabetización académica*. Fondo de Cultura Económica.
- Fraenkel, J. R. & Wallen, N. E. (2000). *How to design and evaluate research in education*. (4th. Ed). Boston: McGraw-Hill.
- Iglesias, P y De Micheli, A. (2009). La incorporación de la escritura a la enseñanza de Biología en primer año de la Universidad. *Obtáculos y estrategias*. Recuperado desde:
http://www.filo.unt.edu.ar/jorn_unesco/cd/PAN%201.2.CARLINO.pdf

- Lucca, N., & Berríos, R. (2009). *Investigación cualitativa: fundamentos, diseño y estrategias*. Hato Rey: Publicaciones Puertorriqueñas.
- Lucero, I. & Meza, S. (2009) Validación de Instrumentos para medir conocimientos. Recuperado desde: <http://fundaeoe.com/curriculum/otro.pdf>
- Morales, P. (2007). Análisis de ítems en las pruebas objetivas.
- Nunnally, J (1987). Teoría Psicométrica. México, editorial Trillas.
- Punch, K. F. (1999). *Introduction to social research: Quantitative & qualitative approaches*. London: Sage.
- Rincke, K. (2006). Learning Physics is like learning a language? Recuperado desde: http://www.physik.uni-kassel.de/did1/Rincke/Download/rincke_esera_2006.pdf.
- Rodríguez, D. (2003). *Medición, "assessment" y evaluación del aprovechamiento académico*. Séptima edición. Publicaciones Puertorriqueñas, Editores. Hato Rey P.R.
- Seidman, I. (1998). *Interviewing as qualitative research: A guide for researchers in education and the social sciences*. New York: Teachers College Press.
- Stake, R. (2000). Case study. In N. Denzin & Y. Lincoln (Ed), *Handbook of qualitative research* (2nd Ed). (pp. 435-454). London: Sage.
- Vera, L. (2004). *Medición, Assessment y evaluación del aprendizaje*. Publicaciones Puertorriqueña editores, Hato Rey, P.R.
- Verdejo, A. y Medina, M. (2008). *Evaluación del aprendizaje Estudiantil*. Cuarta

Edición Isla negra

editores, San Juan, P.R.

Wolcott, H. (1994). *Transforming qualitative data: Description, analysis, and interpretation*. Thousand Oaks, CA: Sage.

Zhang, B. (2008). Application of unidimensional ítems response models to tests with ítems sensitive to secondary dimensions. *The journal of experimental Education*, 77(2). Pp. 147-166.