

VALIDACIÓN DE PRUEBAS DE RENDIMIENTO ACADÉMICO

ARTURO BARRAZA MACÍAS

ISBN: 978-607-98598-0-0



9 786079 859800

VALIDACIÓN DE PRUEBAS DE RENDIMIENTO ACADÉMICO

ARTURO BARRAZA MACÍAS

Primera edición: abril de 2019

Editado en México

ISBN: 978-607-98598-0-0

Editor:

Universidad Pedagógica de Durango.

Respaldo Institucional:

Red Durango de Investigadores Educativos A.C.

Instituto Universitario Anglo Español

Corrector de estilo:

Gonzalo Arreola Medina

Este libro no puede ser impreso, ni reproducido total o parcialmente por ningún otro medio sin la autorización por escrito de los editores

CONTENIDO

INTRODUCCIÓN	6
CAPITULO UNO	9
RENDIMIENTO ACADÉMICO	
<i>Antecedentes: una caracterización sucinta de lo investigado sobre el rendimiento académico</i>	9
<i>Conceptualización del rendimiento académico</i>	13
<i>Pruebas de rendimiento académico</i>	22
<i>Directrices básicas para la elaboración de pruebas de rendimiento académico</i>	24
CAPITULO DOS	30
VALIDACIÓN DE PRUEBAS DE RENDIMIENTO ACADÉMICO	
<i>Confiabilidad</i>	31
<i>Valdez de contenido</i>	37
<i>Análisis de reactivos</i>	47
<i>Índice de dificultad</i>	49
<i>Índice de discriminación</i>	53
<i>El coeficiente de correlación biserial-puntual</i>	57
<i>Índice basado en las proporciones de aciertos</i>	62

<i>Inclusión vs. Exclusión de ítems de una prueba de rendimiento académico</i>	63
ANEXO UNO OPERACIONALIZACIÓN VS. ESPECIFICACIÓN	67
ANEXO DOS CORRECCIÓN PARA LAS RESPUESTAS POR ADIVINANZAS	69
REFERENCIAS	71

INTRODUCCIÓN

La Universidad Pedagógica de Durango ofrece, entre sus programas formativos de posgrado, el Doctorado en Ciencias para el Aprendizaje; este doctorado, y las investigaciones que se realizan en su interior, han puesto a la variable rendimiento académico en el centro de las discusiones metodológicas institucionales

Algunos investigadores han optado por elegir trabajar con el rendimiento académico auto-informado, pero han reconocido las limitaciones que esta toma de decisiones acarrea. Ante esto se ha decidido trabajar con pruebas de rendimiento académico, sin embargo, las decisiones sobre como validar estas pruebas han generado algunas controversias entre el colegio doctoral. Algunos solo aceptan el análisis de reactivos mientras que otros quieren utilizar todos los procedimientos usados en la validación de otros instrumentos de investigación, incluido entre ellos el análisis factorial.

Para superar estas controversias, pero sobre todo, para ofrecer información pertinente para la validación de estas pruebas fue que se inició la redacción del presente libro. En su camino pude constatar que el problema no era privativo de nuestra institución y de nuestro estado, sino que era un problema compartido por la mayor parte de los investigadores del campo educativo, salvo pequeñas células

de expertos en estadística trabajando en institutos o centros de evaluación que se encuentran focalizados en la elaboración de pruebas estandarizadas.

Bajo esta ecología epistémico-metodológica, signada por imprecisiones y vacíos, es que se ofrece el presente libro a la comunidad de investigadores y a los estudiantes de posgrado, maestría o doctorado, interesados en tomar al rendimiento académico como una de sus variables de estudio en la elaboración de sus tesis.

Para su presentación el libro está organizado en dos capítulos. El capítulo uno denominado “Rendimiento académico” y el capítulo dos intitulado “Validación de pruebas de rendimiento académico”.

En el primero se expone el marco prescriptivo y conceptual que circunscribe y le otorga un sentido referencial al contenido central del presente libro a través de cuatro indicadores que funcionan como subtítulos: a) Antecedentes: una caracterización sucinta de lo investigado sobre el rendimiento académico, b) Conceptualización del rendimiento académico, c) Pruebas de rendimiento académico, y d) Directrices básicas para la elaboración de pruebas de rendimiento académico.

En el capítulo dos se abordan la Confiabilidad, la Validez de contenido y el Análisis de reactivos como ejes analíticos centrales, valga la redundancia, que permiten determinar las propiedades psicométrica de una prueba de rendimiento académico y la inclusión o exclusión de ítems o reactivos en la misma.

Finalmente se integran dos anexos que buscan complementar algunos argumentos esbozados al interior del libro. Cierro esta introducción esperando que su contenido sea de utilidad para los potenciales lectores.

CAPITULO UNO

RENDIMIENTO ACADÉMICO

Antecedentes: una caracterización sucinta de lo investigado sobre el rendimiento académico.

El rendimiento académico es una variable abordada en múltiples y variadas investigaciones; hecho que se refleja en la cantidad de estudios realizados y en la diversidad de variables con la que se le relaciona.

A este respecto Google Académico reporta 236,000 resultados; si esta misma consulta se realiza en buscadores más especializados se observa que DIALNET informa de 3,706 documentos alojados en su portal, mientras que REDALyC reporta 7,424 documentos ubicados en su repositorio digital. En una búsqueda selectiva, con un criterio temporal, Google Académico reporta 6,990 resultados para el año 2018. Estas cifras, sin lugar a dudas, son ilustrativas de la presencia de esta variable en la investigación que se realiza actualmente. Cabe mencionar que todas estas consultas fueron realizadas el 28 de diciembre de 2018.

Con relación a la diversidad de variables con la que se le relaciona, la revisión de literatura muestra que lo mismo se le estudia con relación al uso problemático del

internet (Marín, Carballo & Coloma-Carmona, 2018) que a la actividad física (Manzano-Carrasco, López-Serrano, Suárez-Manzano & Ruiz-Ariza, 2018), sin olvidar variables didácticas, como lo es el aprendizaje basado en problemas (Hincapié, Ramos & Chrino-Barceló, 2018), o psicológicas, como lo son los estados emocionales (Pulido & Herrera, 2018).

De igual manera se reportan estudios que relacionan el rendimiento académico con los deberes escolares (Valle, et al. 2015), la implicación familiar percibida (Álvarez et al. 2015), los horarios laborales de los progenitores (Cladellas, Castelló, Muntada & Martin, 2017), la inteligencia emocional (Castaño & Páez, 2015), el engagement (Gómez et al. 2015), las estrategias de aprendizaje (Roux & Anzures, 2015), la lectura por placer (Dezcallar, Clariana, Cladellas, Badia & Gotzens, 2014), el consumo de alcohol (Soliz, Mena & Lara, 2017), la autoestima (Serrano, Mérida & Taberner, 2016), el nivel socioeconómico (Korzeniowski, Cupani, Ison & Difabio, 2016), y el tiempo dedicado a la televisión (Badia, Clariana, Gotzens, Cladellas & Dezcallar, 2015), por mencionar solamente algunas, las cuales representan una muestra ilustrativa de la heurística que acompaña a los estudios sobre el rendimiento académico.

No obstante esta fuerte presencia, en relación a la cantidad y diversidad de estudios, un análisis más fino de las investigaciones realizadas conduce a

reconocer la problemática inherente al estudio de esta variable que se refleja en tres indicadores estructurales:

a) No se esboza un posicionamiento teórico al respecto de esta variable.

Algunos autores se concentran solamente en definirla (González & Vera, 2018; Gómez & Merino, 2018; Horna, 2018; Quintana & Numa, 2018; Sizalima-Cuenca, Vásquez-Mena, Maldonado-Vélez & Ávila-Sánchez 2017); aunque la gran mayoría de ellos ni siquiera la definen (Arias & Campana, 2018; Asenjo & Hernández, 2018; Bernal, Lamos, Vargas, Camargo & Sanchez, 2018; Chacaltana, Valderrama & Chacaltana, 2017; Chavera, Cachicatari, Valdivia & Berrios, 2014; Díaz & Flores, 2018; Hincapié et al. 2018; Marín et al. 2018; Manzano-Carrasco et a. 2018; Regueiro et al. 2018; Rico & Sánchez, 2018; Sarrazola-Moncada et al. 2016; Sifuentes, Pesantes, Rojas, Falcón & Manes, 2017).

b) No se utilizan instrumentos para recolectar la información referida al rendimiento académico, en su defecto, se utilizan centralmente las calificaciones, u otros indicadores que consideren conveniente los investigadores, obtenidas de manera institucional (Álvarez et al. 2015; Bernal et al. 2018; Chiecher, Elisondo, Paoloni & Donolo, 2018; Díaz & Flores, 2018; Estrada, 2018; Gómez et al. 2015; González & Vera, 2018; Goñi, Ros & Fernández-Lasarte, 2018; Horna, 2018; Iniguez-Monroy,

Aguilar-Salinas, De Las Fuentes-Lara & Rodriguez-Gonzalez, 2017; Regueiro et al. 2018; Sanz, Fernández-Martínez, Espada & Orgilés, 2018; Serrano et al. 2016; Trelles, Alvarado & Montánchez, 2018); en algunas ocasiones las calificaciones o el promedio son informados directamente por los propios alumnos (Del Rosal, Moreno-Manso & Bermejo, 2018; Marin et al. 2018; Roux & Anzures, 2015).

- c) Son escasos los trabajos que utilizan instrumentos para medir el rendimiento académico y su denominación varía de un autor a otro, p. ej. prueba (Sifuentes, et al. 2017), test de conocimientos (Arias & Campana, 2018; Hincapié et al. 2018), prueba pedagógica (Chávez, Reyes, Mons, González & Mata, 2018), prueba objetiva (Huerta & Pantoja, 2016), tests de aprovechamiento escolar (Korzeniowski et al. 2016), y prueba de rendimiento académico (Castillo, 2018), por mencionar solamente algunas de estas denominaciones; pero más allá de esta poca presencia, y la variabilidad en su denominación, el dato más ilustrativo es el hecho de que estos pocos instrumentos no son sujetos a un proceso de validación serio, o no se informa el haberlo hecho, salvo los casos de: a) Hincapié et al. (2018) y Huerta y Pantoja (2016) que realizan consulta a expertos, b) Castillo (2018) que desarrolla una consulta a expertos y obtiene el nivel de confiabilidad, y c) Korzeniowski, et al. (2016) que informan los niveles de confiabilidad de su batería de pruebas.

Una mención aparte, de esta caracterización sucinta de lo investigado sobre el rendimiento académico, es la presencia de algunos estudios instrumentales enfocados a informar las propiedades psicométricas de algunas pruebas desarrolladas, p. ej. González, Álvarez y Nashiki (2015) analizan los índices de discriminación, de facilidad y del comportamiento de las opciones de respuesta de una prueba diseñada para medir el rendimiento académico en un curso de psicología evolutiva de la adultez y vejez de una licenciatura en psicología, y Sánchez y Reyes (2015) validan una prueba de comprensión numérica para alumnos de primaria a través de la obtención de la confiabilidad y del análisis de los índices de discriminación y de dificultad de cada reactivo.

En esta revisión, cabe aclarar, no se tomaron en cuenta aquellos estudios orientados a determinados procesos psicológicos, p. ej. Las habilidades del pensamiento (Tapia & Luna, 2010) por considerar que propiamente no son pruebas de rendimiento académico.

Conceptualización del rendimiento académico.

Para entender que es el rendimiento académico es necesario hacer una primera distinción entre dos conceptos: aprendizaje y rendimiento académico (Figura 1).

El aprendizaje es una variable teórica, esto es, para definirla hacemos uso de diversas teorías elaboradas bajo la égida del método hipotético deductivo. Bajo esta lógica se puede definir al aprendizaje desde una teoría constructivista, conductista o cognoscitivista, por mencionar las más comunes.

Un ejemplo nos puede ilustrar al respecto: Aprendizaje significativo es el proceso a través del cual una nueva información (un nuevo conocimiento) se relaciona de manera no arbitraria y sustantiva (no-litera) con la estructura cognitiva de la persona que aprende (Ausubel, 1976),

La importancia de la teoría no solo se refleja en la definición que se proporcione de la variable sino en la forma en que se mide para una indagación empírica.

Siguiendo con el ejemplo ya mencionado es menester primero hacer una acotación para proceder a pensar en cómo se mediría dicha variable: cuando se habla de conocimiento o información se hace referencia a un contenido conceptual y cuando se hace referencia a la estructura cognitiva se hace alusión a una estructura conceptual.

Una vez hecha esta acotación cabría preguntarse ¿cómo se mediría el aprendizaje logrado por un alumno desde esta teoría? La respuesta es clara: primero se tendría que indagar que conceptos ha asimilado el alumno, en términos ausubelianos y, sobre todo, como los ha organizado en su estructura cognitiva

vías las relaciones de coordinación, supraordinación y subordinación que configuran una estructura cognitiva de índole conceptual. Bajo esta lógica la mejor manera de medir el aprendizaje sería preguntando definiciones y elaborando mapas conceptuales.

Si decidimos usar otra definición diferente, proveniente de otra teoría, las implicaciones metodológicas para la indagación empírica serían diferentes, y así sucesivamente por cada teoría.

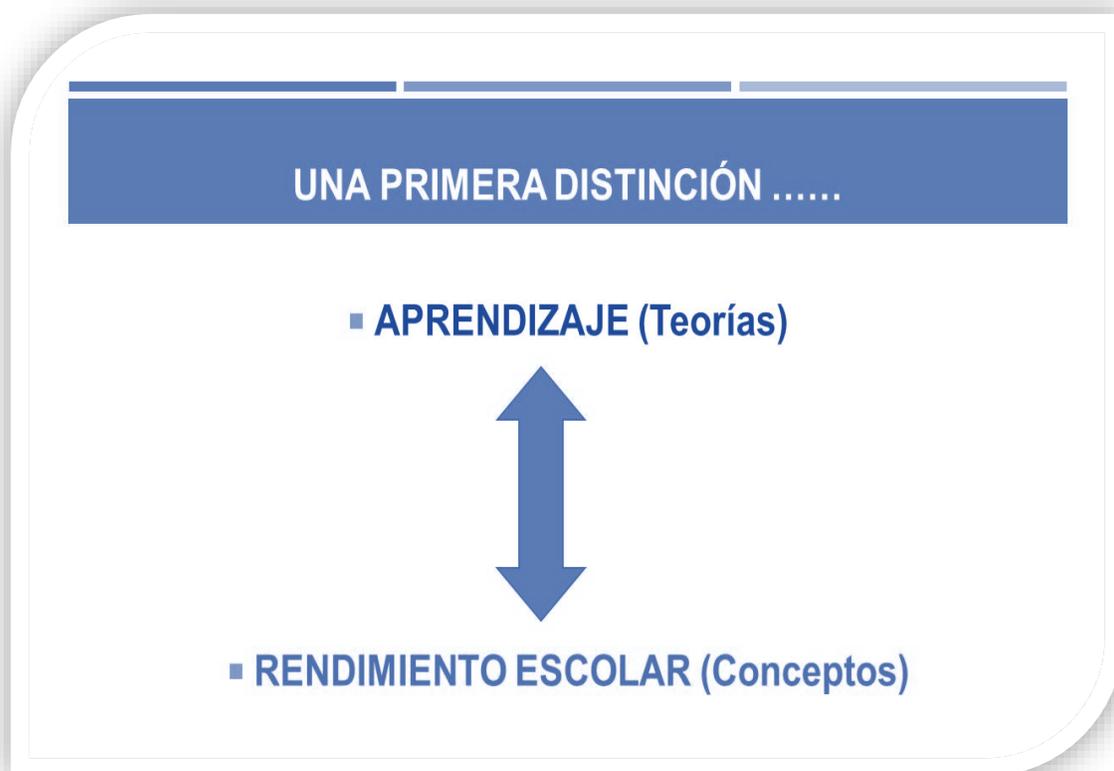
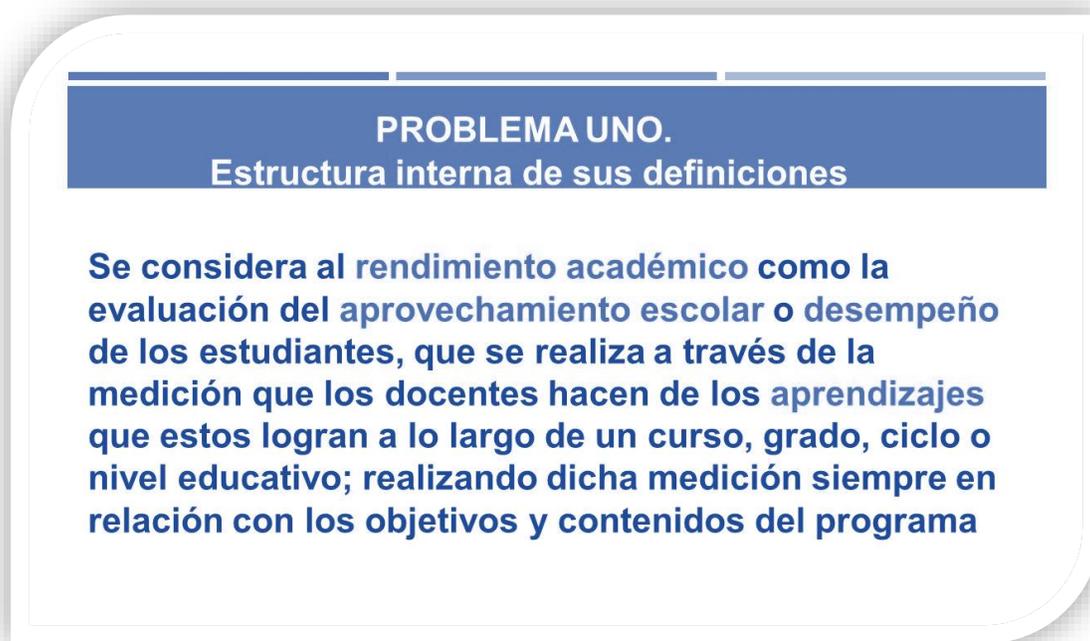


Figura 1. Distinción entre las variables: aprendizaje y rendimiento académico

Por su parte la variable rendimiento académico es una variable a-teórica con un contenido empírico derivado de una conceptualización que, desafortunadamente, suele presentar la mayoría de las veces problemas de delimitación o contenido.

Los problemas para realizar una adecuada conceptualización de esta variable pueden ser de dos tipos: las inconsistencias internas de la definición que se utilice y las relaciones no suficientemente clarificadas con otros términos semejantes.



PROBLEMA UNO.
Estructura interna de sus definiciones

Se considera al rendimiento académico como la evaluación del aprovechamiento escolar o desempeño de los estudiantes, que se realiza a través de la medición que los docentes hacen de los aprendizajes que estos logran a lo largo de un curso, grado, ciclo o nivel educativo; realizando dicha medición siempre en relación con los objetivos y contenidos del programa

Figura 2. Ejemplo del problema tipo uno al intentar conceptualizar la variable rendimiento académico.

Un ejemplo del problema tipo uno se presenta en la Figura dos. Como se puede observar en esa definición los autores (Camarena, Chávez & Gómez, citados en

Morales, Morales & Holguín, 2016, p. 3) juegan con los términos: rendimiento académico, aprovechamiento escolar, desempeño y aprendizaje, lo que sin lugar a dudas le resta claridad al contenido de la definición.

Por su parte la Figura 3 ilustra el problema tipo dos donde diferentes términos se integran al mismo campo semántico, sin claridad de sus límites, e inclusive se usan algunos de ellos como sinónimos (Morales et al. 2016, pp. 4 y 1).

PROBLEMA DOS:
Relaciones de sinonimia con otros términos

El rendimiento escolar traducido a un número también se asocia con términos como: acreditar, medir, valorar, evaluar, entre otros.

En ocasiones se le denomina como aptitud escolar, desempeño académico o rendimiento escolar.

Generalmente las diferencias entre términos sólo se explican por cuestiones semánticas, ya que generalmente, en los textos, la vida escolar y la experiencia docente, son utilizadas como sinónimos.

Figura 3. Ejemplos del problema tipo dos al intentar conceptualizar la variable rendimiento académico.

Más allá de esta problemática, cuando se define al rendimiento académico es menester reconocer que es una construcción social derivada fundamentalmente de imperativos económicos centrados en la búsqueda de la eficacia (Figura 4). A este respecto me permito realizar dos acotaciones:

- a) Creo más adecuado utilizar el término eficacia, que el término eficiencia utilizado por Morales et al. (2016), ya que el término eficacia alude al logro de objetivos y esta característica identitaria permite comprender el proceso de construcción que permea al término rendimiento y lo caracteriza como un término centrado en el producto comportamental de una intervención.

- b) La construcción social se concreta en tres ámbitos empíricos: político, curricular y escolar. En cada uno de ellos, de manera secuencial e interdependiente, se realiza una centración en el producto comportamental o en el logro de objetivos, creando una narrativa de corte economicista para entender la actuación de los agentes educativos.

El rendimiento escolar es una construcción social a partir de las teorías de la economía.



Figura 4. Ámbitos de concreción a través del cual el término rendimiento académico se significa.

En términos metodológicos se pueden destacar sus tres características identitarias (Figura 5): a) es una variable esencialmente empírica (es una variable con un alto grado de especificidad que se circunscribe a una situación concreta y puede ser fácilmente observable), b) es una variable con una finalidad indagatoria de tipo pragmática y utilitarista que permite determinar el resultado de una actuación, y c) es una variable situacional que depende del marco curricular (planes y programas de estudio) que establece los objetivos a lograr.

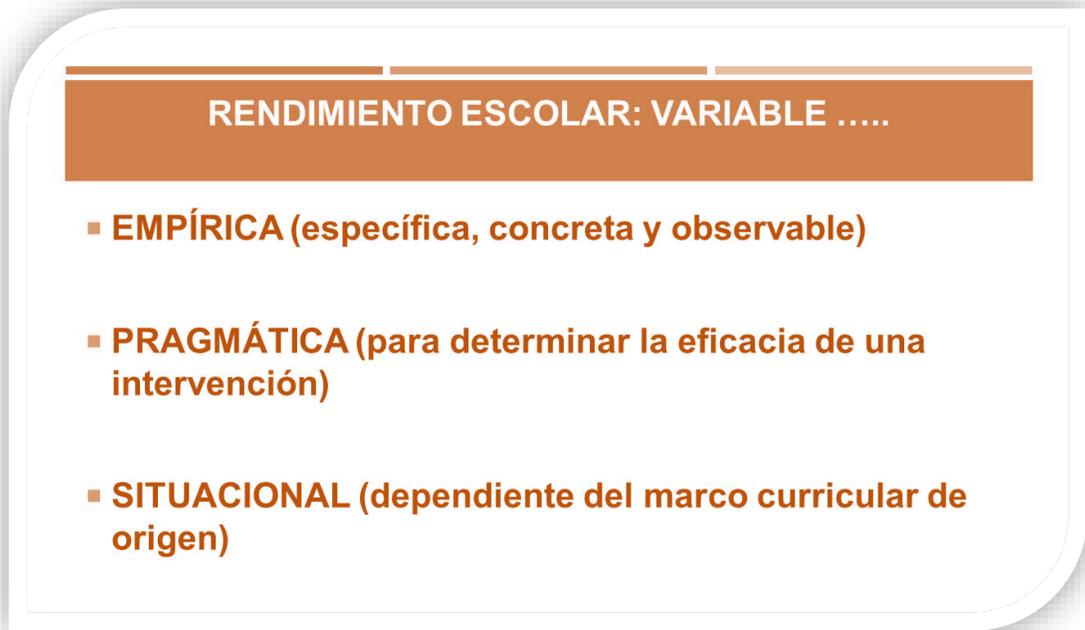


Figura 5. Características metodológicas de la variable rendimiento académico

Estas características conceptuales del rendimiento académico tienen implicaciones importantes a la hora de elaborar un informe final de investigación, sea éste una tesis o un artículo de investigación; la más importante es que en el informe final de investigación no se puede teorizar sobre el rendimiento académico ante la ausencia de teorías al respecto; reconozco algunos esfuerzos de construir modelos teóricos, vía análisis de factores o dimensiones presentes en la literatura al respecto (p. ej. Garbanzo, 2007; Navarro & Blandón, 2017), pero aún estos planteamientos distan mucho de ser una auténtica teoría.

No obstante, estas limitaciones no justifican aquellas investigaciones donde no se define al rendimiento académico (vid supra), sino que considero necesario definir esta variable en el informe de investigación. Para esto se pueden seguir dos caminos:

- Se presentan varias definiciones sobre el rendimiento académico y después el investigador o evaluador tiene que seleccionar la que considere más conveniente para su investigación y justificar su elección.
- Elaborar un constructo (Kerlinger & Lee, 2001) sobre el rendimiento académico, esto es, elaborar su propia definición.

En este segundo caso, que implica mayor trabajo sin duda, recomiendo utilizar la siguiente ruta metodológica, basada esencialmente en la propuesta de Goode y Hatt (1998) para una reconceptualización:

1. Listar todos los términos centrales contenidos en las definiciones existentes
2. Analizar los términos listados y determinar aquellos que realmente pueden ser parte de nuestra definición al aportar algo sustancial y propio a nuestro concepto.
3. Acudir a investigaciones precedentes y analizar las definiciones proporcionadas por los autores y contrastarlas con los términos obtenidos en el segundo paso.

4. Identificar otras variables o términos que se asocian al de rendimiento académico y que, con su propia definición, ayudan a circunscribir el contenido de la nuestra. Mediante un análisis de oposición entre los diferentes términos determinar los elementos que si son propios de nuestra definición.
5. Distinguir el nivel inmediato superior (relación de supraordinación: concepto de mayor jerarquía donde se ubica el nuestro) e inferior (relación de subordinación: conceptos que se derivan del nuestro) de generalización, o de índole conceptual, con los que se relaciona nuestro concepto.

El producto de este análisis será una definición más completa y consistente para el informe final de investigación ya que se filtraron los elementos accidentales de nuestra definición y se recuperaron los elementos esenciales que dan un sentido identitario a nuestra variable.

Pruebas de rendimiento académico.

Los instrumentos que miden el rendimiento académico suelen ser denominados de diferente manera (vid supra), sin embargo, por mi formación pedagógica y mi ascendencia en literatura de habla hispana, prefiero el término prueba al de test (aunque signifiquen lo mismo), y de los diferentes términos que le acompañan (Figura 6) prefiero usar el término “pruebas de rendimiento académico” que es

más fiel a su origen economicista que los términos aprovechamiento o desempeño.

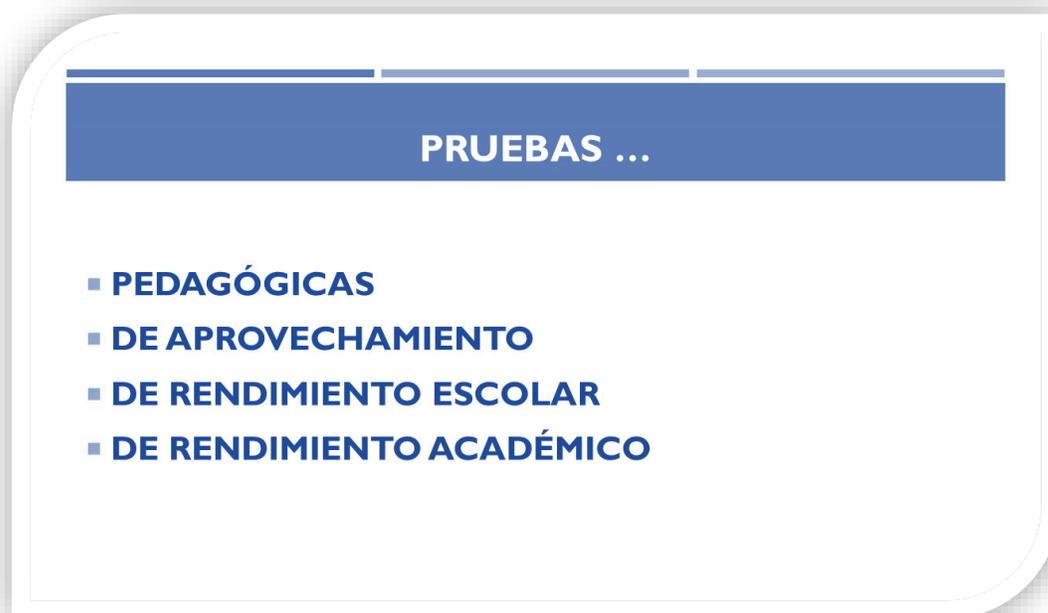


Figura 6. Diferentes denominaciones para los instrumentos que miden el rendimiento académico

Una vez clarificado el término a utilizar es necesario definir qué se entiende por pruebas de rendimiento académico; para esto retomo una definición proporcionada por Lamas (2015) sobre que es evaluación y agregando algunas palabras, no substanciales en términos semánticos pero si gramaticales, me permito definir las pruebas de rendimiento académico en los término marcados en la Figura 7.

No obstante, es menester reconocer que esta definición es genérica y tendría que ajustarse terminológicamente en función del marco curricular que circunscriba el proceso de medición del rendimiento académico.

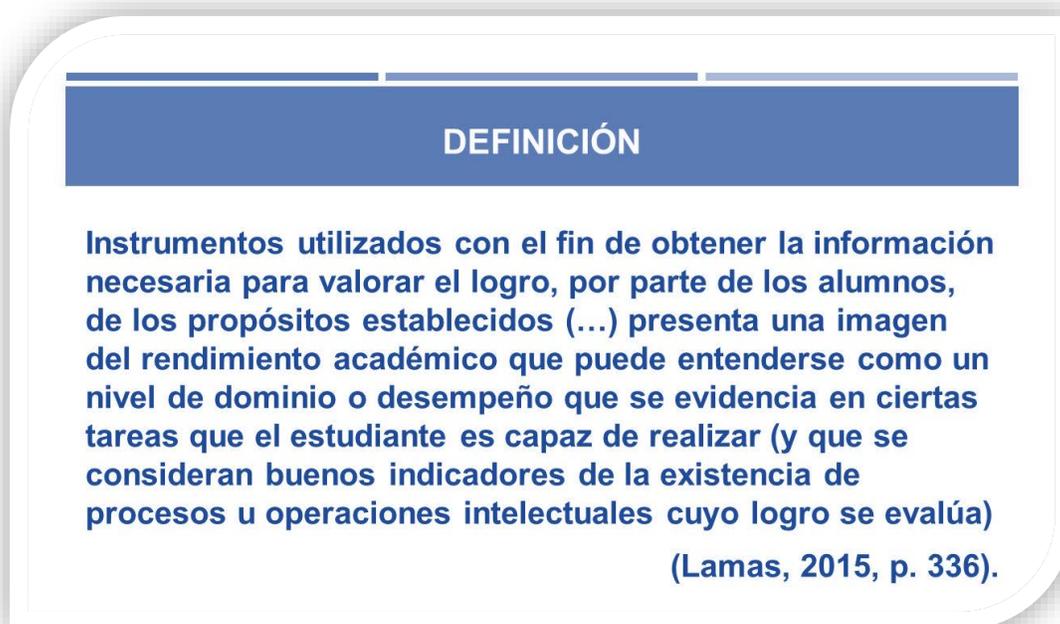


Figura 7. Definición de pruebas de rendimiento académico.

Directrices básicas para la elaboración de pruebas de rendimiento académico.

“La construcción de buenos instrumentos de medición requiere de dos etapas: la elaboración propiamente dicha (...) y el análisis posterior de la misma a partir de los resultados observados una vez que fue administrada a un conjunto de sujetos” (Félix, Galiber & Aguerri, 1999, p.105).

En el presente libro la atención está centrada en la segunda etapa y no es el objetivo del presente libro mostrar el proceso de elaboración de una prueba de rendimiento académico por lo que solo me concentraré en plantear unas directrices básicas a seguir en su proceso de elaboración.

1.- A diferencia de otros instrumentos utilizados en la investigación, p. ej. escalas, inventarios, índices, etc., las pruebas de rendimiento académico no requieren realizar la fase metodológica de la operacionalización (la justificación se puede leer en el anexo 1), en contraparte se suelen utilizar algunos recursos gráficos de organización como lo serían las tablas de especificación donde centralmente se especifican los ítems que corresponden a cada objetivo, propósito o competencia.

2.- Hay dos formas de construir pruebas de rendimiento académico: la relacionada con el criterio y la relacionada con la norma (Brown, 1980); particularmente prefiero los términos la centrada en los objetivos y la centrada en el contenido de aprendizaje. La primera de ellas, que es la de interés para el presente libro, consiste “definir los objetivos para un curso o una unidad dada y, luego, redactar los reactivos de modo que se logren medir esos objetivos” (Brown, 1980, p. 305).

3.- La versión inicial de la prueba siempre debe contener un número mayor de ítems al que se tenga pensado que debe tener la versión final de la prueba. Si se

desea una prueba de 20 ítems es recomendable hacer una primera versión con 30 ítems ya que esto permitirá eliminar aquellos que presenten problemas en el proceso de validación, o en su defecto seleccionar a los que presenten mejores propiedades psicométricas.

4.- Los tipos de preguntas que se pueden utilizar son: a) de respuesta abierta: respuesta breve o completar frases, b) de seleccionar la respuesta: de elección múltiple, elección alternativa, falsa y verdadera, falsa y verdadera múltiple, emparejamiento, elección múltiple compleja y dependientes de un contexto (Moreno, Martínez & Muñiz, 2004). La terminología y clasificaciones de los tipos de pregunta pueden variar aunque esencialmente corresponden a los mismos tipos (vid Muñiz, 2003)

5.- El número de ítems por objetivo, propósito o competencia debe ser el mismo ya que de lo contrario habrá sobrerrepresentación de algunos objetivos en detrimento de otros; la excepción a esta recomendación es cuando deliberadamente se han ponderado de manera diferenciada los diferentes objetivos y esta ponderación se vería reflejada en el número de ítems. En este segundo caso hay que hacer evidente el criterio utilizado en la ponderación diferenciada.

6.-Es recomendable ordenar los ítems o reactivos en la prueba en función de su nivel de dificultad; se debe iniciar con los ítems con menores niveles de dificultad

(Anastasi & Urbina, 1998); estos reactivos se consideran de calentamiento y “se utilizan al comienzo de una prueba para hacer que los alumnos se introduzcan en el examen y tengan confianza” (Brown, 1980, p. 311)

7.- En el caso de los ítems o reactivos de elección múltiple, que son los más comunes, las respuestas erróneas u opciones distractoras se deben elegir entre las respuestas incorrectas más frecuentes; en ese sentido, es recomendable hacer un ejercicio previo donde una muestra de estudiantes respondan al ítem, sin opciones de respuesta, solamente por recuerdo o evocación, eligiéndose posteriormente las respuestas erróneas que se presentaron con mayor frecuencia para ser las opciones distractoras (Brown, 1980).

8.- El tiempo aproximado de respuesta a un ítem o reactivo de elección múltiple es de un minuto (Brown, 1980), por lo que se sugiere que las pruebas tengan una extensión máxima de 40 ítems para que en su respuesta el alumno ocupe entre 40 y 50 minutos. En caso de que la necesidad del investigador lo lleve a elaborar una prueba con mayor extensión es recomendable que su aplicación se de en dos etapas.

9.- Es conveniente, en los ítems que presenten respuestas de opción múltiple, realizar un análisis de las alternativas de respuesta mediante una tabla de contingencia que muestre el número de personas que respondieron cada opción

de respuesta, tanto del grupo alto, como del grupo bajo (Cohen & Swedlik, 2004).

A continuación presento un ejemplo.

En una aplicación de una prueba a 40 estudiantes, una vez evaluadas las pruebas se ordenaron en forma descendente según el puntaje obtenido y se tomó el 25% de los alumnos ubicados al inicio (calificaciones más altas) y se les identificó como el grupo alto, posteriormente se tomó al 25% de los alumnos que fueron los últimos en la lista (calificaciones más bajas) y se les denominó el grupo bajo. Los resultados mostrados en el ítem o reactivo uno se presentan en la Tabla 1:

Tabla 1
Proporción de respuesta en los dos grupos de análisis

Reactivo 1	Opciones			
	a) respuesta incorrecta	b) respuesta incorrecta	c) respuesta correcta	d) respuesta incorrecta
Grupo alto	0	1	8	1
Grupo bajo	0	3	2	5

Estos resultados son típicos y adecuados ya que fueron más los miembros del grupo alto, que los miembros del grupo bajo, los que respondieron correctamente este ítem; sin embargo este tipo de análisis pueden reportar algunos problemas a los que hay que prestar atención:

- Cuando una cantidad alta de miembros del grupo alto eligen una opción distractora en particular.

- Cuando todos los miembros del grupo alto respondieron la respuesta correcta, mientras que los miembros del grupo bajo respondieron de manera dispersa entre las opciones distractoras.
- Cuando la cantidad de miembros del grupo alto que respondieron la opción correcta es baja.
- Cuando la cantidad de miembros del grupo alto, que respondieron la respuesta correcta, es menor a los miembros del grupo bajo que respondieron correctamente.

En cualquiera de estos casos, u otros similares, el investigador debe revisar las repuestas ofrecidas al ítem, sobre todo los distractores. Si hay necesidad habría que realizar algunas entrevistas fast track con algunos alumnos para obtener mayor información y saber que sucedió y si hay necesidad de reformular la redacción de las opciones distractoras o eliminar algunas.

10.- Si la aplicación de la prueba de rendimiento escolar se realiza con fines investigativo es recomendable que el investigador realice una corrección a la calificación final en función de las respuestas por adivinanzas (Anexo dos).

CAPITULO DOS VALIDACIÓN DE PRUEBAS DE RENDIMIENTO ACADÉMICO

La validación de una prueba de rendimiento académico es el proceso que, a través de diferentes procedimientos estadísticos realizados a una base de datos construida a partir de la aplicación del instrumento en cuestión, determina las propiedades psicométricas de su proceso de medición y de los puntajes obtenidos en su aplicación.

De las diferentes propiedades psicométricas que se pueden estudiar cuando se valida un instrumento de investigación, en el caso de las pruebas de rendimiento académico se deben estudiar la confiabilidad y la validez (Figura 8), pero de las diferentes evidencias de validez que se pueden recolectar (Elosua, 2003) considero necesario solamente centrar la atención en la validez de contenido y la validez basada en la estructura interna, específicamente en el análisis de ítems; ambos tipos pueden ser considerados evidencias internas o centradas en los ítems.

El análisis de la confiabilidad y de las evidencias de validez, basadas en el contenido y la estructura interna, conducirán a identificar los ítems buenos de una prueba. “Un buen reactivo de prueba es confiable y válido. Además un buen

reactivo de prueba ayuda a discriminar a quienes responden la prueba; asimismo es uno que responderán en forma correcta quienes obtienen puntuaciones altas en la prueba en conjunto” (Cohen & Swedlik, 2004, p. 240).

Confiabilidad.

La primera propiedad psicométrica a estudiar es la confiabilidad y bajo este rubro se agrupan todo un conjunto de técnicas utilizadas por los investigadores para valorar “la consistencia del proceso de medición o de los resultados” (Bonilla, 2006: 1).

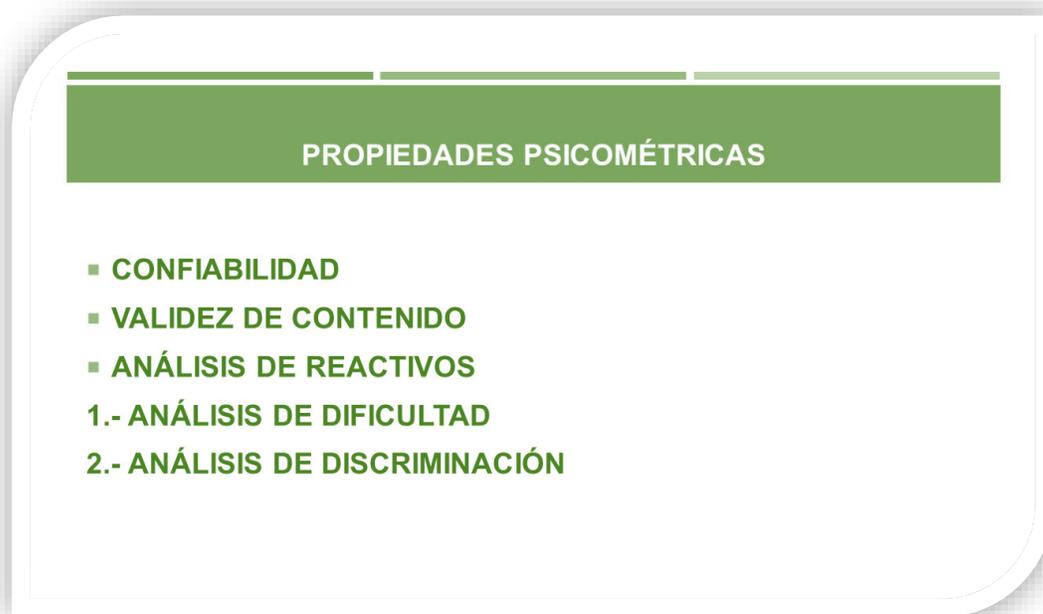


Figura 8. Propiedades psicométricas a indagar en la validación de pruebas de rendimiento académico.

Las diferentes técnicas o procedimientos que se pueden utilizar para la obtención de la confiabilidad son:

- La correlación entre dos formas paralelas de la prueba.
- La correlación entre dos mitades aleatorias de la prueba corregida mediante la fórmula de Spearman-Brown.
- La correlación entre dos aplicaciones de la misma prueba
- La consistencia interna de una prueba.

En el caso de las pruebas de rendimiento académico se recomienda utilizar la correlación entre dos mitades, corregida mediante la fórmula de Spearman-Brown, y la de Kuderr Richardson o el alfa de Cronbach (Figura 9) ya que como, “el examen típico de rendimiento escolar tiene sólo una forma y se administrará solo una vez, no hay más que dos métodos para obtener una estimación de su confiabilidad: el de mitades y el de consistencia interna” (Brown, 1980, pp. 333-334).

La confiabilidad por mitades ofrece una medida de consistencia interna a partir de la correlación entre una y otra mitad y se utiliza la fórmula de Spearman-Brown para su corrección, mientras que el coeficiente alfa de Cronbach ofrece una medida de consistencia interna a partir de las correlaciones item-item.

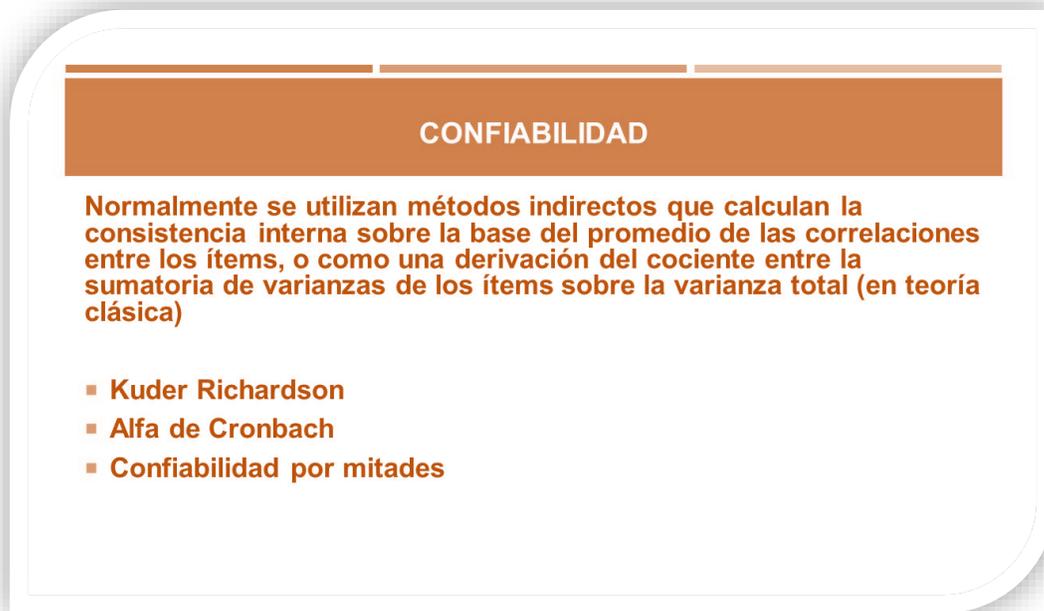


Figura 9. Tipos de confiabilidad a utilizar en el proceso de validación de una prueba de rendimiento académico.

Personalmente prefiero los procedimientos de consistencia interna que brindan correlaciones ítem-ítem, ya que la confiabilidad por mitades puede ser buena, o no, dependiendo de la estructura de la prueba.

Aclaremos esta última afirmación. Si una prueba aborda solo un contenido de aprendizaje o inclusive un solo aspecto de ese contenido (p. ej. suma de centenas o solamente su algoritmo) la confiabilidad por mitades puede ser una buena opción ya que no importa los ítems que queden en una u otra mitad se está evaluando el mismo contenido o el mismo aspecto del contenido.

Pero qué pasa si la prueba evalúa diferentes contenidos o inclusive contenidos de diferentes asignaturas (p. ej. reactivos que abordan contenidos de historia, de matemáticas o de español); en este caso se corre el riesgo de que las dos mitades no tengan ítems del mismo nivel de dificultad, de discriminación, de la misma asignatura o del mismo contenido y eso afecte la respuesta de los alumnos y, por consecuencia, afecte la confiabilidad de la prueba.

No obstante esta aclaración, y la toma de postura personal al respecto, considero que queda a criterio del evaluador utilizar uno o los dos procedimientos que se pueden trabajar mediante el programa Paquete Estadístico para las Ciencias Sociales (SPSS, por sus siglas en inglés): alfa de Cronbach y Confiabilidad por mitades.

A continuación menciono la ruta para la obtención de la confiabilidad con ambos procedimientos:

1.- En primer lugar se construye la base de datos en el programa SPSS. La codificación es dicotómica (acierto-error) sin importar si se utilizaron reactivos de opción múltiple con tres o cuatro opciones de respuesta. Se recomienda utilizar en la codificación numérica el cero para error y el uno para acierto.

2.- Una vez capturados los datos se abre la ventana de *Analizar*, se posiciona el cursor en *Escala* y se da un clic en *Análisis de Fiabilidad*.

3.- En la ventana que se despliega aparecen dos columnas: en la primera aparecen todos los ítems o reactivos (variables) de la prueba y la segunda aparece vacía.

4.- Usando el cursor y la función seleccionar se van trasladando cada uno de los ítems o reactivos a la columna derecha usando la flecha que está en medio de las dos columnas.

5.- La primera columna, ubicada del lado izquierdo, tiene en la parte inferior otra ventana abierta que al lado izquierdo tiene el título de *Modelo*; en esta ventana aparece por default el modelo *Alfa*. Para el caso del alfa de Cronbach dejamos esta opción intacta, si se desea utilizar la confiabilidad por mitades se despliega el menú de esta opción y se selecciona confiabilidad por mitades.

6.- Una vez trasladados los ítems o reactivos que componen el cuestionario a la segunda ventana, ubicada al lado derecho, y habiendo constatado que el modelo que aparece seleccionado por default es el *Alfa* (o se cambió el modelo en caso de desear obtener la confiabilidad por mitades) se le da clic a al botón *Aceptar* ubicado, junto con otros, en el lado interior del cuadro.

7.- El SPSS brindará el coeficiente de confiabilidad en un valor que va de 0 a 1. Este coeficiente puede ser interpretado, igual que con cualquier instrumento de investigación, en función de escala de valores ya propuestas por diferentes autores; Un ejemplo es la escala propuesta por Murphy y Davishofer (en Hogan, 2004); estos autores ofrecen la siguiente escala de valores:

- Alrededor de .90 es un nivel elevado de confiabilidad.
- La confiabilidad de .80 o superior puede ser considerada como moderada.
- Alrededor de .70 se considera baja.
- Inferior a .60 indica una confiabilidad inaceptablemente baja.

A este análisis de la confiabilidad, de carácter general o total de la prueba de rendimiento académico, se puede agregar otro análisis consistente en obtener el nivel de confiabilidad de la prueba en caso de eliminar un ítem. Para realizar este análisis seguimos los primeros cinco pasos ya mencionados y posteriormente hacemos los siguientes:

6.- Una vez trasladados los ítems o reactivos que componen el cuestionario a la segunda ventana, ubicada al lado derecho, y habiendo constatado que el modelo que aparece seleccionado por default es el *Alfa*, se le da click al botón del lado derecho que dice *Estadísticos* y en el menú que se despliega se selecciona la

opción *Escala si se elimina un elemento*, se da click al botón de continuar y regresamos a la ventana anterior y se le da click al botón *Aceptar* ubicado, junto con otros, del lado derecho.

7.- El SPSS brindará, aparte del coeficiente de confiabilidad en un valor que va de 0 a 1, una tabla con cuatro columnas; de estas columnas vamos a prestar atención a la última que tiene por encabezado “*confiabilidad si se elimina un ítem* (Tabla 2)”. En esta columna se revisa el coeficiente reportado en el entendido de que si el valor de alfa sube mínimo una décima o cinco centésimas es un ítem que está afectando negativamente a la confiabilidad general de la prueba y se debe considerar la opción de eliminarlo.

Tabla 2
Confiabilidad en caso de eliminar un ítem

	Media de la escala si se elimina el elemento	Varianza de la escala si se elimina el elemento	Correlación elemento-total corregida	Alfa de Cronbach si se elimina el elemento
Ítem 1	34.07	147.543	.606	.922
Ítem 2	32.73	145.747	.708	.917
Ítem 3	32.71	147.493	.688	.918
Ítem 4	32.80	145.296	.765	.914
Ítem 6	31.65	154.451	.597	.922

Validez de contenido.

De acuerdo a Messick (1995) las diferentes evidencias de la validez se pueden agrupar en cinco tipos:

- 1) Evidencia basada en el contenido
- 2) Evidencia basada en el proceso de respuesta:
- 3) Evidencia basada en la estructura interna:
- 4) Evidencia basada en otras variables
- 5) Evidencia basada en las consecuencias de la medición

En el caso del presente libro la atención está puesta exclusivamente en la validez de contenido ya que se coincide con Brown (1980) cuando afirma que para las pruebas de rendimiento académico “la validez de contenido es el procedimiento primordial para el establecimiento de la validez” (p. 334).

El procedimiento para la obtención de evidencias de validez, basadas en el contenido, agrupa dos tipos de estudios considerados suplementarios: unos encaminados a evaluar las relaciones entre el contenido de aprendizaje a evaluar y el contenido de la prueba de rendimiento académico, y otros dirigidos a valorar los factores contextuales internos y externos que añaden varianza no deseada a los resultados de la prueba (Elosua, 2003).

La atención de este apartado está centrada solamente en el primer tipo de estudios que buscan obtener evidencias para determinar si los ítems o reactivos de la prueba de rendimiento académico representan adecuadamente el contenido de aprendizaje que se pretende evaluar.

Para lograr este objetivo tradicionalmente se plantean cuatro estrategias: la fundamentación en una teoría, la consulta a expertos, el análisis de errores y la supervisión de los resultados totales (Anastasi & Urbina, 1998). De estas cuatro estrategias son las dos primeras las que pueden ayudar de mejor manera a la validación de las pruebas de rendimiento académico (Figura 10),

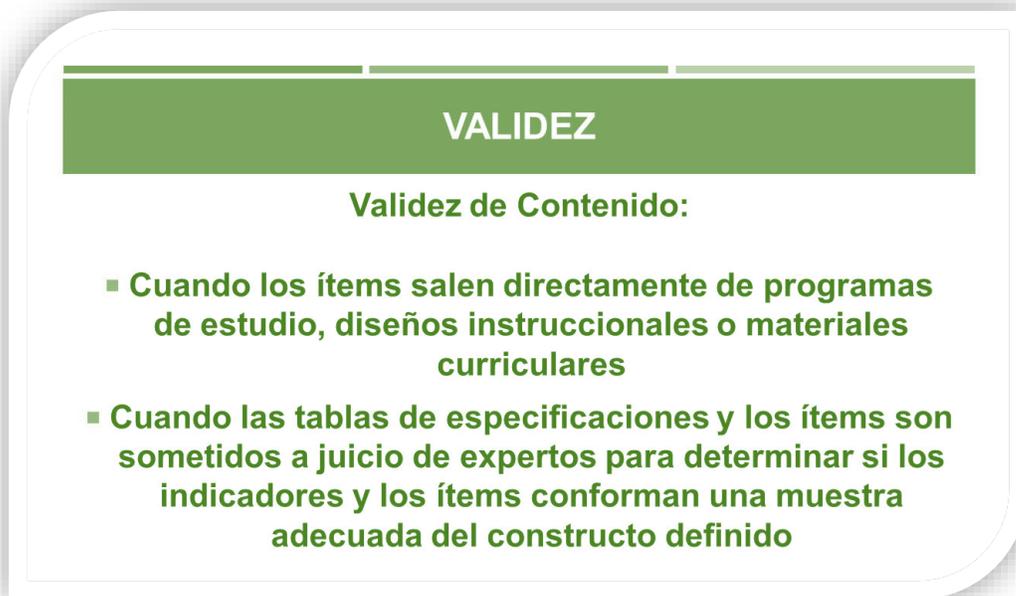


Figura 10. Estrategias para obtener evidencias de validez de contenido de una prueba de rendimiento académico

En el primer caso se considera, de manera general, que la evidencia basada en el contenido la realiza el mismo investigador apoyándose en la teoría y los supuestos que guían la selección de los ítems (Briones, 2001); en ese sentido “el dominio de contenido de una variable, normalmente, es definido o establecido por la literatura

(teorías y estudios antecedentes) (Hernández, Fernández & Baptista, 2006, p. 279). En el caso de las pruebas de rendimiento académico el dominio del contenido de aprendizaje es establecido por el marco curricular de referencia (plan y programas de estudio).

El segundo caso consiste en preguntar a uno o varios expertos si los reactivos evalúan lo que uno quiere que evalúen (Salkind, 1999); en el caso de las pruebas de rendimiento académico se les pregunta a los expertos si el ítem o reactivo evalúa el objetivo, propósito o competencia que pretende evaluar.

Estas dos estrategias de validación, la fundamentación teórica (en nuestro caso curricular) y la consulta a expertos, son consideradas por algunos autores (Briones, 2001; Hernández et. al. 2006; Muñiz, 2003), como complementarias o simultáneas mientras que otros autores (Gregory, 2001). sostienen que la consulta a expertos solamente se usa en caso de que la variable de estudio sea definida de manera aproximada o imprecisa por la ausencia de una buena fundamentación teórica (en nuestro caso una imprecisa relación con el marco curricular).

Como lo mencioné en el año 2007 (p. 11)

Particularmente, en el caso de esta estrategia, coincido plenamente con Gregory (2001), por dos razones: primero, porque es necesario recordar

que el aspecto central de este tipo de evidencia es la relevancia, entendida ésta como la relación de la variable, y los ítems, con la teoría utilizada (Pardo, 2006), y segundo, porque de esta manera se supera la crítica que considera a este tipo de evidencia como algo subjetivo (González, 2006)

Una vez hecha esta aclaración procedo a explicar cada una de estas estrategias. En el primer caso los ítems o reactivos salen directamente del marco curricular (plan y programas de estudio) u otros materiales curriculares como libros de texto, ficheros didácticos o guías del maestro (Figura 11).

Pero ¿qué quiere decir “salen directamente”? esto quiere decir que en la redacción del ítem se toma literalmente el dato, el concepto o la proposición obtenida de alguna de esas fuentes y solamente se redacta en función del tipo de reactivo que se desee utilizar, y en el caso de procedimientos se toman los ejemplos obtenidos en algún material curricular y se sustituyen solamente los datos por otros del mismo nivel de complejidad.

Para justificar que los ítems o reactivos se fundamentan en el marco curricular es necesario elaborar una tabla con dos columnas: en la columna izquierda se ponen los ítems que conforman la prueba y en la columna derecha la fuente curricular en que se fundamenta. Esta tabla se presenta en el informe de validación.

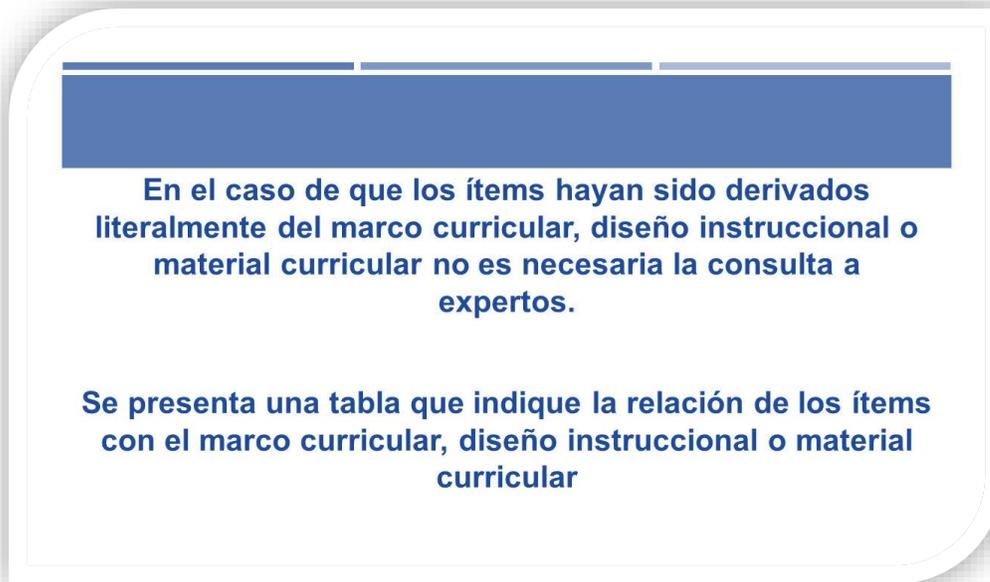


Figura 11. Estrategia de fundamentación en el marco curricular

En el caso de la consulta a expertos se elabora la “tabla de especificaciones” (Rivas, González & Delgado, 2010, p. 286) y cada uno de los expertos establece, de manera independiente, y sin conocer las respuestas de los otros expertos, la correspondencia Ítem – Objetivo.

Es menester hacer una aclaración al respecto: uso de manera genérica y como ejemplo la relación ítem – Objetivo, pero en realidad el segundo término puede ser propósito, competencia, aprendizaje esperado, u otro término que utilice el marco curricular que se tome de referencia.

Una vez hecha esta aclaración continuo: a cada experto se le presenta una tabla con las celdas vacías (en las filas irían los ítems y en las columnas los objetivos) y se le pide que ubique cada ítem en un objetivo; para esto deben utilizar los números 1 y 0, ya sea que el ítem evaluado corresponda o no a cada objetivo, respectivamente.

En el informe de validación se presenta la tabla de especificaciones y los resultados obtenidos en cada ítem (Figura 12); la literatura al respecto no marca un baremo para la toma de decisiones por lo que creo conveniente establecer el siguiente, en términos de porcentaje que indique el acuerdo positivo entre expertos, por lo que dependiendo del número de expertos invitados se tendría que transformar sus resultados en porcentaje:

- Menor al 50% se elimina el ítem o reactivo.
- De 50% al 66% se considera un ítem con validez de contenido débil.
- De 67% al 83% se considera un ítem con validez de contenido aceptable.
- De 84% al 100% se considera un ítem con validez de contenido fuerte.

Este procedimiento, anteriormente descrito, es para obtener el nivel de validez de contenido de un ítem o reactivo pero si se desea obtener un nivel de validez de contenido de toda la prueba de rendimiento académico se obtiene la media general y se interpreta con el mismo baremo.

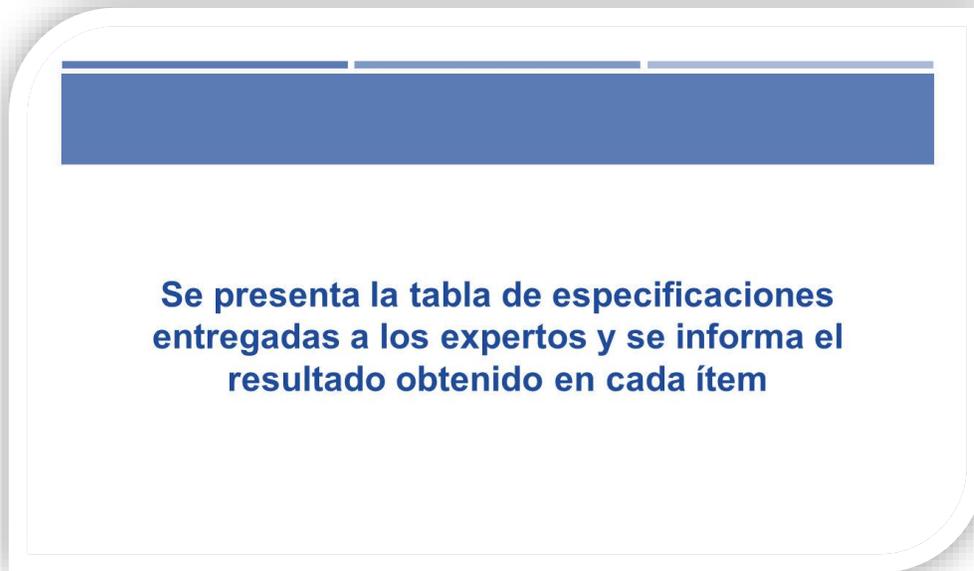


Figura 12. Presentación de los resultados de la consulta a expertos

Cabe hacer mención que el baremo propuesto es una traslación del que propuse para los instrumentos de investigación en general (Barraza, 2007).

Aparte de este procedimiento, algunos de mis alumnos han utilizado en la consulta a expertos una variante del procedimiento que recomiendo para cuestionarios en general (Barraza, 2007). A continuación transcribo el procedimiento haciendo los ajustes necesarios para el caso de las pruebas de rendimiento académico.

En primera instancia se elabora la prueba de rendimiento académico con las indicaciones y la escala que se muestra en el siguiente ejemplo de formato a entregar a los expertos:

Le pedimos su colaboración para establecer la validez de contenido de la siguiente prueba de rendimiento académico., en caso de aceptar le sugerimos tomar en cuenta las siguientes instrucciones:

- *Lea cuidadosamente las indicaciones de cada una de las secciones de esta prueba de rendimiento académico.*
- *A continuación señale qué tan bien considera usted que el ítem o reactivo sea parte del objetivo propuesto.*
- *Valore cada ítem o reactivo de manera individual comparando el contenido del ítem con el objetivo propuesto.*
- *Para evaluar cada ítem o reactivo utilice la siguiente escala de valores:*
 - *El ítem no pertenece al objetivo planteado (NP).*
 - *El ítem probablemente no pertenece al objetivo planteado (PNP).*
 - *El ítem probablemente si pertenece al objetivo planteado (PSP).*
 - *El ítem si pertenece al objetivo planteado (SP).*

En el siguiente listado se ofrecen una serie ítems o reactivos que pedimos evalúe de manera individual para determinar si pertenece o no al objetivo planteado; señale con una X la opción que considere más adecuada.

Objetivo: XXXXXXXXXXXXXXXX

Ítems	No pertenece	probablemente no pertenece	probablemente si pertenece	Si pertenece
X				
XX				
XXX				
XXXX				

Una vez elaborado este formato se les solicita su llenado a los expertos y una vez que este respondidos se codifican los resultados con los siguientes valores: 0 (NP), 1 (PNP), 2 (PSP) y 3 (SP).

Se obtiene la media de cada ítem y en caso de que la misma sea menor a 1.5 el ítem se elimina. Con los ítems restantes se obtiene una media general de la prueba de rendimiento académico o de cada uno de los objetivos planteados y el resultado se interpreta con la siguiente escala de valores:

- De 1.6 a 2.0 se considera que presenta una validez débil.
- De 2.1 a 2.5 se considera que presenta una validez aceptable
- De 2.6 en adelante se considera que presenta una validez fuerte.

La cantidad de expertos a consultar en uno u otro procedimiento es la misma. Particularmente recomiendo como número óptimo cinco expertos y como mínimo tres expertos. Esta cantidad puede variar pero se recomienda mantener un número no menor y no exceder de 10 expertos.

Análisis de reactivos.

De los diferentes tipos de evidencias de validez, mencionados previamente (vid supra), la atención del presente apartado se centra en la evidencia basada en la estructura interna. Este tipo de evidencias se obtienen a través de diferentes procedimientos, algunos centrados en la dimensionalidad de la prueba (Análisis Factorial Exploratorio o Confirmatorio) o en el análisis de reactivos (Análisis de Consistencia interna, de Grupos Contrastados, etc.).

No considero relevante o necesario realizar análisis factoriales en las pruebas de rendimiento académico ya que la dimensionalidad está dada por el marco curricular de referencia y tratar de confirmarla o identificar otra a partir de las respuestas de los evaluados crearía dificultades para conciliar esos resultados con el marco curricular.

Por su parte el análisis de reactivos lo considero relevante e importante para la validación de una prueba y en eso coincido con Kerlinger y Lee (2002) cuando

afirman que el análisis de reactivos puede ayudar a incrementar las evidencias de validez y confiabilidad de las pruebas de rendimiento académico (Figura 13).

En este tipo de análisis se evalúa cada reactivo en lo particular para decidir o no su inclusión final en la prueba.

ANÁLISIS DE REACTIVOS

- **Kerlinger y Lee (2002) afirman que para incrementar la confiabilidad y validez de una prueba es posible ejecutar un conjunto de procedimientos estadísticos llamados análisis de reactivos.**
- **Consiste en evaluar cada reactivo de forma separada para determinar si el reactivo es bueno o pobre**

Figura 13. El análisis de reactivos como fuente de evidencias de validez

Para el análisis de reactivos se proponen en este libro tres procedimientos: análisis de dificultad, análisis de discriminación y correlación biserial.

Índice de dificultad.

El análisis de dificultad nos lleva a determinar el nivel de facilidad/dificultad que representa la respuesta del ítem para la población a evaluar (Figura 14). En ese sentido lo ideal es sostener dentro de una prueba un ítem que tenga una dificultad moderada ya que en nada nos sirve un ítem que responda correctamente toda la población o que no lo responda correctamente nadie.

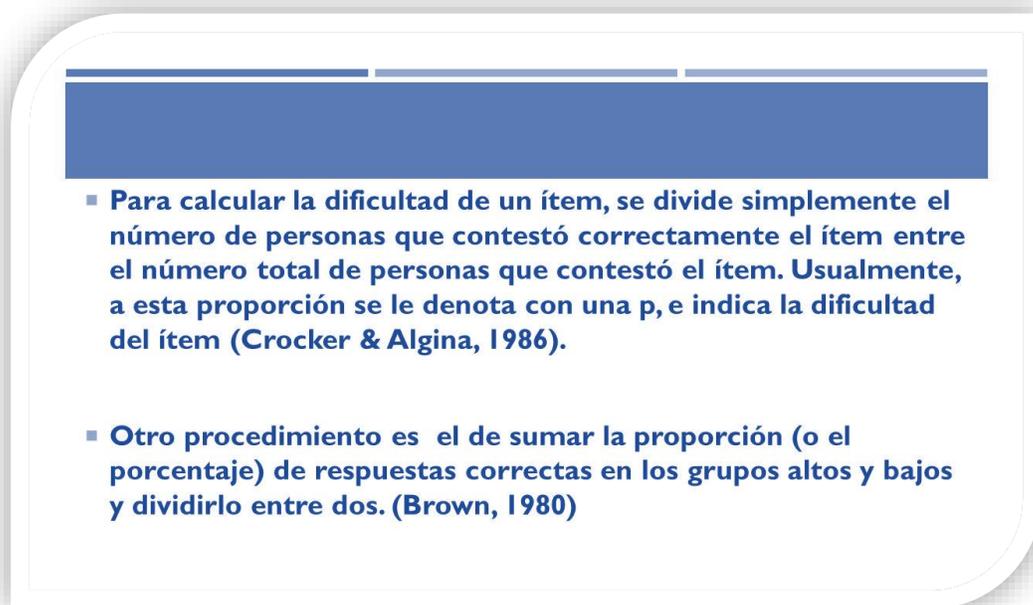
ÍNDICE DE DIFICULTAD.

La dificultad de un ítem se entiende como la proporción de personas que responden correctamente un reactivo de una prueba. Entre mayor sea esta proporción, menor será su dificultad. Lo que quiere decir que se trata de una relación inversa: a mayor dificultad del ítem, menor será su índice (Wood, 1960; en Backhoff, Larrazolo & Rosas, 2000, p. 14)

Figura 14. El índice de dificultad de un reactivo

Para calcular el índice de dificultad existen diferentes procedimientos. En la figura 15 se presentan dos de ellos. Cualquiera puede ser utilizado sin mayores problemas y solamente se requieren cálculos manuales.

El índice de dificultad presenta valores que van de 0 a 1. Cuando los estudiantes evaluados responden de manera correcta casi todos los reactivos o ítems de una prueba, los valores que se obtienen como el índice de dificultad serán muy cercanos a 1. Por el contrario, una prueba que contenga reactivos muy difíciles generará que los estudiantes evaluados respondan de manera incorrecta a casi todos los reactivos o ítems y los valores del índice de dificultad serán muy cercanos a 0.



- **Para calcular la dificultad de un ítem, se divide simplemente el número de personas que contestó correctamente el ítem entre el número total de personas que contestó el ítem. Usualmente, a esta proporción se le denota con una p_i e indica la dificultad del ítem (Crocker & Algina, 1986).**
- **Otro procedimiento es el de sumar la proporción (o el porcentaje) de respuestas correctas en los grupos altos y bajos y dividirlo entre dos. (Brown, 1980)**

Figura 15. Procedimientos para obtener el índice de dificultad de un reactivo

Una vez obtenido el índice de dificultad se interpreta con alguno de los criterios establecidos en la Figura 16, a los que denomino de interpretación directa ya que se asume la inclusión o exclusión de un ítem a partir de su adecuado nivel de dificultad. A diferencia de la propuesta de Backhoff et al (2000) que proponen la configuración de una prueba a partir de diferentes niveles de dificultad (Figura 17), siguiendo la lógica de una distribución normal, yo prefiero asumir la interpretación directa.

INTERPRETACIÓN DIRECTA

De acuerdo al manual del EXHCOBA, el nivel medio de dificultad del examen debe oscilar entre 0.5 y 0.6. (Backhoff, Larrazolo & Rosas, 2000, p. 14)

La mayoría de los creadores de pruebas concuerdan en que los mejores reactivos, en términos de dificultad, son aquellos con valores entre .5 y .7 (Kerlinger & Lee, 2002).

Una regla empírica útil es la de utilizar reactivos a las que respondan correctamente de 50 al 60% de los estudiantes (Brown, 1980, p. 351)

Figura 16. Criterios de interpretación directa para determinar la inclusión de un ítem en una prueba a partir de su nivel de dificultad

Estos criterios para la interpretación directa pueden ser ampliados hacia abajo: a 0.4 (40%) o hasta 0.3 (30%), pudiendo quedar intervalos de 0.3 (30%) a 0.7 (70%) o de 0.4 (40%) a 0.6 (60%); Cohen & Swedlik (2004) aceptan la posibilidad de

manejar un intervalo entre 0.3 (30%) a .08 (80%). Sin embargo esta decisión se tomaría siempre y cuando los ítems o reactivos presenten problemas para lograr el nivel de dificultad óptimo (0.5 ó 50%) y estando consciente el investigador que con esta acción disminuye la calidad de los reactivos de la prueba y por ende la adecuada medición del rendimiento académico.

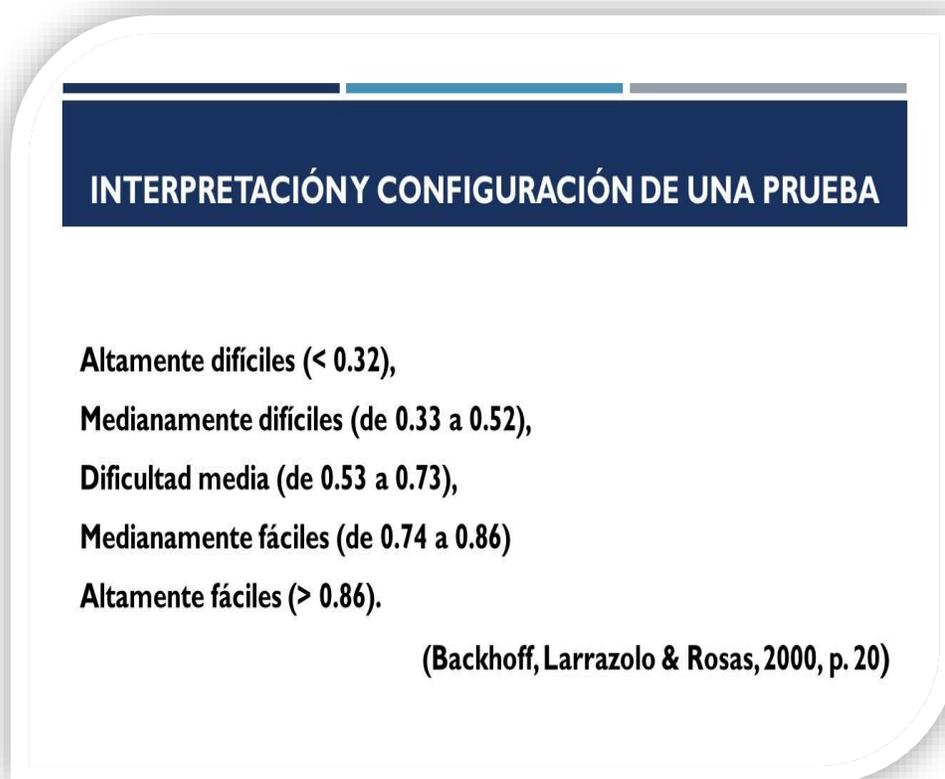


Figura 17. Baremo para la interpretación y configuración de una prueba a partir del nivel de dificultad de los ítems

Como complemento se puede obtener un índice promedio de dificultad de la prueba. “Esto se logra sumando los índice de dificultad del reactivo, para todos los

reactivos de la prueba, y dividiéndolos entre el número total de reactivos en la prueba". (Cohen & Swedlik, 2004, pp. 241-242).

Índice de discriminación.

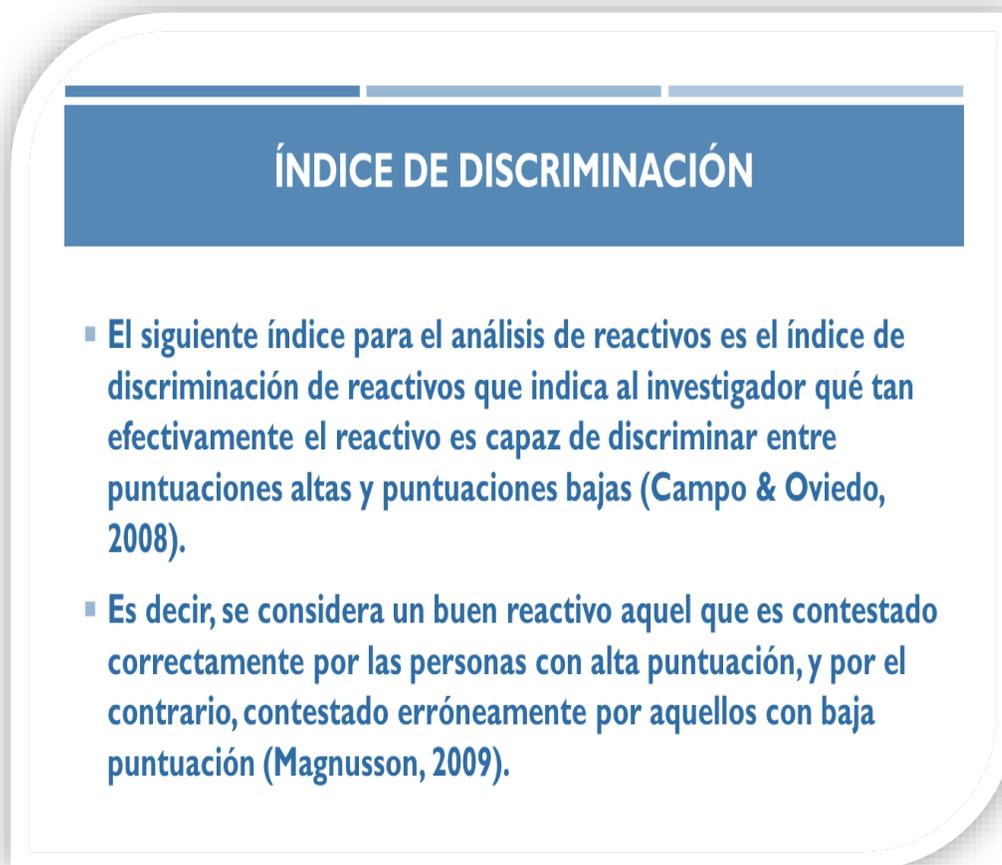
El análisis de discriminación es el procedimiento a través del cual se revisa cada ítem para determinar si el responderlo correcta o incorrectamente permite diferenciar entre el grupo con alto rendimiento y el grupo con bajo rendimiento. En el entendido de que un buen reactivo es aquel que es respondido correctamente por los alumnos con una alta puntuación o calificación, y por el contrario, es respondido erróneamente por aquellos alumnos con baja puntuación o calificación (Magnusson, 2009).

El resultado del análisis se concentra en un valor numérico denominado índice de discriminación (Figura, 18).

Valores de 0.0., 1.0 y -1.0 son extraños, en especial el valor negativo que muestra una discriminación invertida (los participantes de menor puntaje – que saben menos- contestan correctamente). Los otros dos valores son interpretados como de una absoluta discriminación (1.0) o bien una nula discriminación (0.0), en ambos casos el reactivo no cumple la función de diferenciar lo que pretende medir el reactivo (Camacho, 2008b). El valor

esperado del índice de discriminación tiene que ser positivo y muy próximo a uno; a mayor valor, mayor discriminación (Aiken, 2003) (Juárez, 2015, p.35).

El procedimiento para obtener el índice de discriminación se presenta en la Figura 19 y 20, así como la forma de interpretar el resultado obtenido.



ÍNDICE DE DISCRIMINACIÓN

- El siguiente índice para el análisis de reactivos es el índice de discriminación de reactivos que indica al investigador qué tan efectivamente el reactivo es capaz de discriminar entre puntuaciones altas y puntuaciones bajas (Campo & Oviedo, 2008).
- Es decir, se considera un buen reactivo aquel que es contestado correctamente por las personas con alta puntuación, y por el contrario, contestado erróneamente por aquellos con baja puntuación (Magnusson, 2009).

Figura 18. Índice de discriminación

- Para calcular el índice de discriminación de los reactivos el primer paso es ordenar las puntuaciones totales, de la más alta a la más baja. En seguida, el total de participantes se divide en dos grupos: los de alta puntuación y los de baja puntuación.
- El criterio de clasificación de ambos grupos varía según el autor; la mayoría coincide en utilizar el percentil 25 ó 27. Es claro que al considerar los extremos (percentil 25), la probabilidad de encontrar mejores índices de discriminación es mayor. (Caso, Espinoza, Rodríguez y Urias, 2010)

Figura 19. Procedimiento para obtener el Índice de discriminación (primera parte)

- El índice más simple sería el de observar las diferencia en las proporciones o porcentajes: porcentaje de respuestas correctas en el grupo alto menos el porcentaje de respuesta correctas en el grupo bajo.
- Como regla empírica podemos decir que cualquier diferencia de 15 a 20% (0.15-0.20) o mayor indicará una buena discriminación.

(Brown, 1980, p. 352)

Figura 20. Procedimiento para obtener el Índice de discriminación (segunda parte)

Además del procedimiento descrito en las Figuras 19 y 20, existe otro procedimiento y criterio de interpretación que se presentan en las Figuras 21 y 22.

ÍNDICE DE DISCRIMINACIÓN

- ESTE ÍNDICE PERMITE IDENTIFICAR EL PODER DISCRIMINATORIO DE LOS ÍTEMS
- SE IDENTIFICA CON ID.
- EN PRUEBAS DE RENDIMIENTO

$$ID = \frac{A_s - A_I}{n}$$

A_s aciertos grupo superior
 A_I aciertos grupo inferior
 n número de sujetos del subgrupo

- EN ESCALAS DE ACTITUD SE COMPARAN LAS MEDIAS DE LOS DOS GRUPOS

Torrado (s/f), p. 52

Figura 21. Segundo procedimiento para obtener el Índice de discriminación (primera parte)

Tabla I. Poder de discriminación de los reactivos según su valor D

D =	Calidad	Recomendaciones
> 0.39	Excelente	Conservar
0.30 - 0.39	Buena	Posibilidades de mejorar
0.20 - 0.29	Regular	Necesidad de revisar
0.00 - 0.20	Pobre	Descartar o revisar a profundidad
< -0.01	Pésima	Descartar definitivamente

Ebel y Frisbie (1991)

Backhoff, Larrazolo & Rosas, 2000, p. 15

Figura 22. Segundo procedimiento para obtener el Índice de discriminación (segunda parte)

El coeficiente de correlación biserial-puntual.

La obtención del índice de discriminación tiene como principal limitante que se encuentra asociado al índice de dificultad (Anastasi & Urbina, 1998) por lo que se recomienda complementarlo con un coeficiente de correlación. Bajo esta lógica “el índice de discriminación se define como la correlación entre las puntuaciones de los sujetos en el ítem y sus puntuaciones en el test” (Muñiz, 2003, p. 219).

Los coeficientes de correlación que se pueden utilizar son:

- Correlación biserial-puntual (una variable dicotómica y otra continua)
- Correlación biserial (una variable dicotómica de manera no natural y otra continua)
- Coeficiente phi (ambas variables dicotómicas)
- Correlación tetracórica (ambas variables dicotómicas de manera no natural)

“Dado que es habitual que los ítems sean dicotómicos (o se aciertan o se fallan), y el test constituye una medida cuantitativa descrita” (Muñiz, 2003, p. 220) se utiliza normalmente el coeficiente de correlación biserial-puntual

“El coeficiente biserial puntual r_{ix} es el coeficiente de correlación de Pearson entre el ítem i (dicotómico) y el puntaje total en la prueba que lo contiene” (Félix et al.

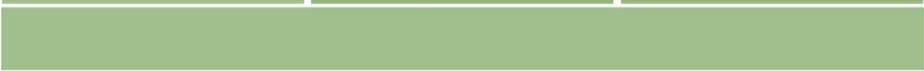
1999, p. 107). La idea detrás de este análisis es que la calificación total de la prueba obtenida por cada alumno evaluado se relaciona con el continuo hipotético del reactivo presentado en términos dicotómicos (Figura 23).

COEFICIENTE DE CORRELACIÓN BISERIAL

Se calcula para determinar el grado en que las competencias que mide el test también las mide el reactivo. El rbis proporciona una estimación de la correlación producto-momento de Pearson entre la calificación total de la prueba y el continuo hipotético del reactivo, cuando éste se dicotomiza en respuestas correctas e incorrectas (Henrysson, 1971).

Figura 23. Coeficiente de correlación biserial

Para interpretar el resultado obtenido generalmente se considera que cuanto más alta es la correlación entre el ítem y la calificación total de la prueba, el ítem representa mejor a la prueba (Félix et al. 1999); sin embargo, también se puede interpretar siguiendo la escala de valores propuesta por Hernández et al. 2006) (Figuras 24 y 25).



Se interpreta por el mismo estándar de Pearson, lo que cambia en el análisis es la consideración del signo ya que el mismo en este coeficiente no se interpreta.

Figura 24. Interpretación del coeficiente de correlación biserial (primera parte)

- 
- **Correlaciones entre $\pm.15$ y $\pm.10$ se consideran como bajas**
 - **Entre $\pm.30$ y $\pm.40$ como moderadas**
 - **Entre $\pm.50$ y $\pm.70$ como moderadamente altas**
 - **Entre $\pm.80$ y $\pm.90$ como altas**
 - **Y más de $\pm.90$ muy altas**

Figura 25. Interpretación del coeficiente de correlación biserial (segunda parte)

A esta forma de interpretación hay que hacerle una acotación en el sentido de que no son recomendables valores muy altos ya que esto afectaría al índice de dificultad, el cual presentaría una gran dispersión (Anastasi & Urbina, 1998)

Para obtener este coeficiente de correlación en el programa SPSS se sigue la siguiente ruta de análisis:

- 1.- En primer lugar se construye la base de datos en el programa SPSS. La codificación del ítem es dicotómica (acierto-error) sin importar si se utilizaron reactivos de opción múltiple con tres o cuatro opciones de respuesta. Se recomienda utilizar en la codificación numérica el cero para error y el uno para acierto.
- 2.- Se integra una nueva variable donde se capturan las calificaciones o puntajes totales obtenidos por cada alumno en la prueba.
- 3.- Una vez capturados los datos se abre la ventana de *Análisis*, se posiciona el cursor en *Correlación* y se da un clic en la opción *bivariada*.
- 4.- En la ventana que se despliega aparecen dos columnas: en la primera aparecen todos los ítems o reactivos (variables) de la prueba, así como la variable

que reporta la calificación o puntaje final obtenido en la prueba y la segunda aparece vacía.

5.- Usando el cursor y la función seleccionar se van trasladando cada uno de los ítems o reactivos a la columna derecha, así como la variable que reporta la calificación o puntaje final obtenido en la prueba, usando la flecha que está en medio de las dos columnas.

6.- En la parte inferior de estas columnas aparecen tres opciones de análisis, estando seleccionada por default "*Pearson*", la cual se deja.

7.- En la fila siguiente, o parte inferior de estas opciones, aparecen dos celdas que tienen por encabezado "*prueba de significación*"; las celdas tienen los encabezados de *bilateral* y *unilateral*. Por default aparece seleccionada la opción de *bilateral* y ésta es la que se deja.

8.- El SPSS brindará el coeficiente de confiabilidad en un valor que va de 0 a 1, pero que solo adquiere sentido si el nivel de significación es menor a .05. Si se cubre el criterio de significación el coeficiente es interpretado como se indica en la figura 25. Hay que tener cuidado de no perderse con los diferentes valores que aporta la tabla que brinda el programa; para esto se toma como referencia que los ítems sean las filas y la columna sea la media general o puntaje total obtenido en

la prueba. Los resultados que quedan en las celdas donde se cruzan son los únicos que interesan.

Índice basado en las proporciones de aciertos.

Otra opción para el análisis de la discriminación del ítem, sin tomar en cuenta la correlación ítem-test mencionado en el apartado anterior, es el índice basado en las proporciones de aciertos. Este es un índice muy elemental pero ofrece información complementaria que ayuda a conocer la capacidad discriminativa de los ítems.

Este índice es la diferencia entre la proporción de alumnos con altas calificaciones que responden correctamente el ítem y la proporción de alumnos con bajas calificaciones que también lo responden correctamente al ítem (Muñiz, 2003). Para la conformación de los grupos con alta y baja calificación se sigue el procedimiento marcado en la Figura 19.

La interpretación de este índice y de su capacidad discriminativa, está en función de lo cerca o lejos que esté de cero; entre más cerca esté del cero el valor obtenido se puede considerar que el ítem tiene un bajo índice y por lo tanto se considera un ítem con nula capacidad discriminativa. Un valor negativo sería

catastrófico ya que implicaría que los alumnos de baja calificación son los que contestan en mayor cantidad el ítem de manera correcta.

Inclusión vs. Exclusión de ítems de una prueba de rendimiento académico.

Los diferentes procedimientos abordados en este capítulo ofrecen información sobre las propiedades psicométricas de la prueba de rendimiento académico, en lo general, o de los ítems que la conforman, en lo particular. Sin embargo hay necesidad de discutir ¿cuándo tomar la decisión de eliminar un ítem de una prueba o modificarlo para aplicarlo en una segunda ocasión y ver su comportamiento psicométrico nuevamente?

La respuesta no es sencilla, de hecho la mayoría de los textos, en lo general, no lo abordan; en lo particular esta situación suele mencionarse en cada análisis, en lo individual, y sugerir la eliminación o modificación de un ítem en función de ese análisis en lo particular, p. ej. el índice de dificultad.

A este respecto considero que se debe tomar una decisión de manera global, tomando como referencia los diferentes procedimientos utilizados; p. ej. en el presente libro se han abordado seis procedimientos (confiabilidad, validez de contenido, índice de dificultad, índice de discriminación, coeficiente de correlación

biserial-puntual y el índice basado en las proporciones de aciertos), por lo que la decisión de eliminar o modificar un ítem debería de estar en función de su comportamiento general.

Con este planteamiento me distancio del posicionamiento purista que exige eliminar algún ítem si tienen problemas en el índice de dificultad o de discriminación, por enunciar los más importantes o los mencionados de manera más recurrente.

Para tomar la decisión de manera global el investigador o evaluador debe establecer una regla de decisión basada en los diferentes procedimientos utilizados; p. ej. dado el caso de que se usaran los seis procedimientos descritos en este capítulo una regla de decisión podría ser la siguiente:

- Eliminar un ítem si presenta problemas en por lo menos tres de los análisis efectuados.
- Modificar un ítem si presenta problemas en uno o dos de los análisis utilizados.

Esta regla de decisión se puede flexibilizar (aceptando que solo se eliminaría en caso de presentar problemas en cuatro de los análisis efectuados) o hacer más severa (aceptando se eliminen ítems en caso de presentar problemas en dos o

más de los análisis realizados) en función de los intereses del propio investigador o evaluador.

Así mismo, se puede modificar esta regla de decisión ponderando de manera diferenciada cada análisis desarrollado, p. ej. de los seis análisis efectuados se le puede dar prioridad al índice de dificultad y establecer una regla de decisión en los siguientes términos:

- Eliminar un ítem si presenta problemas en por lo menos tres de los análisis efectuados, o en dos de ellos, si uno de ellos es el índice de dificultad.
- Modificar un ítem si presenta problemas en uno o dos de los análisis utilizados, siempre y cuando ninguno de ellos sea el índice de dificultad.

En lo personal propondría la siguiente regla de decisión:

- Eliminar un ítem si presenta problemas en la mitad de los procedimientos desarrollados, ya sea que se utilicen seis o cuatro.
- Eliminar un ítem si presenta problemas de manera simultánea en el índice de dificultad y en el coeficiente de correlación biserial-puntual.
- Modificar un ítem si presenta problemas en uno de los procedimientos utilizados, o máximo en dos, siempre y cuando estos dos no sean el índice de dificultad o el coeficiente de correlación biserial-puntual.

En los casos que se menciona la necesidad de modificación implica reformularlo, volver a aplicarlo y nuevamente efectuar el análisis de sus propiedades psicométricas.

En los casos que se indica la eliminación de ítems el investigador debe revisar que todos sus objetivos, propósitos, competencias o aprendizajes esperados queden representados de manera equitativa con los ítems restantes, salvo que deliberadamente se haya ponderado de manera diferenciada cada uno. Si no se cubre este requisito hay que elaborar nuevos ítems y someterlos a validación.

ANEXO UNO OPERACIONALIZACIÓN VS. ESPECIFICACIÓN

La operacionalización es una fase epistémico-metodológica dentro del método hipotético deductivo que pretende derivar de manera lógica, e informada teóricamente, consecuencias observacionales de variables que adquieren su pleno sentido en el seno de una teoría.

En términos metodológicos estas consecuencias observacionales darán lugar a los ítems del instrumento de investigación y para esto se sigue una secuencia metodológica que permite derivar de la variable las dimensiones y de éstas los indicadores, los cuales servirán de base a los ítems del instrumento.

Este proceso se realiza sobre objetos de conocimiento construidos teóricamente, por lo que entre más consolidada este la teoría que lo circunscribe más sencillo y consistente será el proceso de operacionalización.

En el caso de las pruebas de rendimiento académico nos distanciamos de este proceso ya que trabajamos sobre contenidos de aprendizaje y no sobre objetos de conocimiento. Y la pregunta obligada es ¿Cuál es la diferencia?

Cuándo se elabora un plan de estudios los diseñadores se enfrentan a la decisión de elegir qué objetos de conocimiento se integrarán a ese plan de estudios. Una vez decidido sigue el problema de cómo transformarlo en contenido de aprendizaje, en el entendido de que no necesariamente la forma en que se organiza el conocimiento teóricamente o la forma en que fue construido por la comunidad científica es el camino o lógica que sigue un niño en su construcción.

Este proceso, no suficientemente estudiado, se denomina transposición didáctica y su consecución sigue varias líneas de desarrollo mediadas por teorías psicológicas o prácticas socioculturales, por lo que la organización, la secuencia y los desagregados presentes en los contenidos de aprendizaje no necesariamente corresponden a la planteada por la comunidad científica.

Hecha esta aclaración se puede entender por qué no se realiza una operacionalización para la elaboración de una prueba de rendimiento académico; es preferible partir de los objetivos, propósitos o competencias, que al fin de cuentas reflejan la intencionalidad educativa de los diseñadores que han sido sancionadas institucionalmente, y de ellos derivar los reactivos o ítems de la prueba.

Este argumento también justifica el por qué no es recomendable realizar una validación factorial de una prueba de rendimiento académico.

ANEXO DOS CORRECCIÓN PARA LAS RESPUESTAS POR ADIVINANZAS

En la respuesta a los diferentes ítems, independientemente del formato en que se presenten, cabe siempre la posibilidad de que el alumno acierte por adivinanza. Frase como “De tin marin de do pingue” o “ave maría, afíname la puntería” son una clara muestra de que los alumnos muchas veces recurren a la estrategia de seleccionar una respuesta al azar abriendo la posibilidad de atinarle a la respuesta correcta aunque no la sepan.

En el caso que nos interesa, que es la validación de pruebas de rendimiento académico para fines investigativos, creo que es necesario que el evaluador o investigador realice una corrección a la calificación obtenida con el fin de asegurar que la calificación refleje realmente el rendimiento obtenido por el alumno.

La corrección consiste en restar una porción de las respuestas incorrectas a las respuestas correctas mediante la siguiente fórmula (Brown, 1980, p. 330):

$$\text{Calificación correcta} = R - \frac{W}{N - 1}$$

R = número de respuestas correctas

W = número de respuestas incorrectas (no se contabilizan si se dejaron en blanco)

N = número de opciones de respuesta por pregunta.

Un ejemplo para ilustrar el uso de esta fórmula sería el siguiente:

- De una prueba de 40 reactivos el alumno responde correctamente 28 reactivos (R); y de manera incorrecta 8 (W) dejando sin responder cuatro.
- Las preguntas están redactadas con cuatro opciones de respuesta (N).
- La sustitución de la fórmula sería la siguiente:

$$\text{Calificación correcta} = 28 - \frac{8}{4 - 1}$$

- Si redondeamos a tres el resultados de 8 entre 3, el puntaje correcto para este alumno sería 25.

REFERENCIAS

- Álvarez, A., Suárez, N., Tuero, E., Núñez, J.C., Valle, A., & Regueiro, B. (2015). Implicación familiar, autoconcepto del adolescente y rendimiento académico. *European Journal of Investigation in Health, Psychology and Education*, 5(3), 293-311. doi:10.1989/ejihpe.v5i3.133
- Anastasi, A. & Urbina, S. (1998). *Test psicológicos*. México: Prentice Hall.
- Arias, R., & Campana, A. (2018). Gestión pedagógica y su relación con el rendimiento académico del área de comunicación del 5to grado de las I.E.P. del nivel primario distrito de Tambo-La Mar-Ayacucho-2015. *Big Bang Faustiniiano*, 7(3), 17-24. Recuperado de <http://revistas.unjpsc.edu.pe/index.php/BIGBANG/article/view/430/421>
- Asenjo, J.A., & Hernández, R.W. (2018). Estrategias de aprendizaje y rendimiento académico en estudiantes de un instituto superior tecnológico privado. *Revista Ciencia Nor@ndina*, 1(1), 31-37. Recuperado de http://unachinvestiga.edu.pe/ciencianorandina/images/yt-sampleddata/revistas-cientificas/revista-cientifica-vol1/4_estrategias.pdf
- Ausubel, D.P. (1976). *Psicología educativa: un punto de vista cognoscitivo*. México, Trillas.
- Backhoff, E., Larrazolo, N., & Rosas, M. (2000). Nivel de dificultad y poder de discriminación del Examen de Habilidades y Conocimientos Básicos

- (EXHCOBA). *REDIE. Revista Electrónica de Investigación Educativa*, 2(1), 11-28., Recuperado de <http://www.redalyc.org/pdf/155/15502102.pdf>
- Badia, M.M., Clariana, M., Gotzens, C., Cladellas, R., & Dezcallar, T. (2015). Videojuegos, televisión y rendimiento académico en alumnos de primaria. *Pixel-Bit. Revista de Medios y Educación*, 46, 25-38.
doi:10.12795/pixelbit.2015.i46.02
- Barraza, A. (2007). La consulta a expertos como estrategia para la recolección de evidencias de validez basadas en el contenido. *Investigación Educativa Duranguense*, 2(7), 5-14. Recuperado de <http://www.upd.edu.mx/PDF/Revistas/InvestigacionEducativaDuranguense7.pdf>
- Bernal, M.I., Lamos, A.F., Vargas, O.I., Camargo, G.E., & Sánchez, N. (2018). Enfoques de aprendizaje, rendimiento académico y factores relacionados en estudiantes que cursan último año de los programas de la Facultad de Ciencias de la Salud. *Educación Médica*, 311, 1-8.
doi:10.1016/j.edumed.2017.11.008
- Bonilla, V.E. (2006). Confiabilidad. Boletín informativo INEVA en acción, 2(3), 1-3.
Recuperado de <http://ineva.uprrp.edu/boletin/v0002n0003.pdf>
- Briones, G. (2001). *Métodos y técnicas de investigación para las ciencias sociales*. México: Trillas.
- Brown, F.G. (1980). *Principios de la medición en psicología y educación*. México: El manual moderno.

- Campo, A., & Oviedo, H. C. (2008). Propiedades psicométricas de una escala: la consistencia interna. *Revista Salud Pública*, 10(5), 831-839. Recuperado de <https://scielosp.org/pdf/rsap/2008.v10n5/831-839>
- Caso, J., Espinoza, L.A., Contreras, L.A., Rodríguez, J.C., & Urias, E. (2010). *Propiedades psicométricas de la adaptación del Cuestionario de Estrategias de Aprendizaje para adolescentes. UEE RT 10-001*. México: Universidad Autónoma de Baja California. Recuperado de <http://uee.uabc.mx/reportesTecnicos/10-001.php>
- Castaño, J.J., & Páez, M.L. (2015). Inteligencia emocional y rendimiento académico en estudiantes universitarios. *Psicología desde el Caribe*, 32(2), 268-285. doi:10.14482/psdc.32.2.5798
- Castillo, N. (2018). *Programa de Educación Adaptativa en el desarrollo de las competencias matemáticas en los estudiantes. El Porvenir 2017*. (Tesis doctoral inédita). Universidad César vallejo, Perú. Recuperada de <http://repositorio.ucv.edu.pe/handle/UCV/22682>
- Chacaltana, J.A., Valderrama, J., & Chacaltana, A.A. (2017). Asertividad y su relación con el rendimiento académico de los estudiantes de primer año de la Escuela Profesional de Ingeniería Metalúrgica y Materiales de la Universidad Nacional Jorge Basadre Grohmann – Tacna, 2016. *Ciencias*, 1(1), 38-42. Recuperado de <http://revistas.unjbg.edu.pe/index.php/cs/article/view/493/439>

- Chavera, L., Cachicatari, E., Valdivia, C., & Berrios, Y. (2014). Factores intrínsecos e intrínsecos relacionados al rendimiento académico de los estudiantes de Ciencias de la Salud UNJBG. Tacna 2012. *Revista Médica Basadrina*, 8(1): 35 – 41. Recuperado de <http://revistas.unjbg.edu.pe/index.php/rmb/article/view/586/524>
- Chávez, J.L., Reyes, E., Mons, M.A., González, L.E., & Mata, J.A. (2018). Desarrollar la habilidad escribir en inglés con propósitos específicos, en ciencias médicas. *EDUMECENTRO*, 10(1), 156-167. Recuperado de <http://scielo.sld.cu/pdf/edu/v10n1/edu10118.pdf>
- Chiecher, A.C., Elisondo, R.C., Paoloni, P.V., & Donolo, D.S. (2018). Creatividad, género y rendimiento académico en ingresantes de ingeniería. *Revista Iberoamericana de Educación Superior*, 9(24), 138-151. Recuperado de <http://www.scielo.org.mx/pdf/ries/v9n24/2007-2872-ries-9-24-138.pdf>
- Cladellas, R., Castelló, A., Muntada, M.C., & Martín, M.B. (2017). Horarios laborales de los progenitores y su incidencia en el rendimiento académico de alumnos de primaria. Efectos diferenciales del género. *Revista Portuguesa de Educação*, 30(1), 135-155. doi:10.21814/rpe.9512
- Cohen, R.J., & Swedlik, M.E.(2004). *Pruebas y evaluación psicológicas*. México: Mc Graw Hill.
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winsto.

- Del Rosal, I., Moreno-Manso, J.M., & Bermejo, M.L. (2018). Inteligencia emocional y rendimiento académico en futuros maestros de la Universidad de Extremadura. *Profesorado. Revista de Currículum y Formación del Profesorado*, 22(1), 257-275. Recuperado de <https://recyt.fecyt.es/index.php/profesorado/article/view/63644/38702>
- Dezcallar, T., Clariana, M., Cladellas, R., Badia, M., & Gotzens, C. (2014). La lectura por placer: su incidencia en el rendimiento académico, las horas de televisión y las horas de videojuegos. *Ocnos*, 12, 107-116.
doi:10.18239/ocnos_2014.12.05
- Díaz, B.G., & Flores, Y.H. (2018). Factores que influyen en el rendimiento académico de los estudiantes de educación en la Universidad Católica De Trujillo Benedicto XVI de Trujillo, Perú 2017- II. *SEARCHING*, 1(1), 1-13.
Recuperado de <http://revista.uct.edu.pe/index.php/HUMANITIES/article/view/40/27>
- Elosua, P. (2003). Sobre la validez de los test. *Psicothema*, 15(2), 315-321.
Recuperado de <http://www.psicothema.es/pdf/1063.pdf>
- Félix, H., Galiber, M.S., & Aguerri, M.E. (1999). Valoración de los ejercicios en las pruebas de rendimiento escolar. *Educación Matemática*, 11(3), 104-119.
Recuperado de <http://www.revista-educacion-matematica.org.mx/descargas/Vol11/3/10Atorresi.pdf>
- Garbanzo, G.M. (2007). Factores asociados al rendimiento académico en estudiantes universitarios, una reflexión desde la calidad de la educación

superior pública. *Revista Educación*, 31(1), 43-63. Recuperado de <https://revistas.ucr.ac.cr/index.php/educacion/article/view/1252/1315>

Gómez, F.M., & Merino, A.Z. (2018). Estrategias de aprendizaje y rendimiento académico empleadas por estudiantes de liceos de Talca con alto y bajo puntaje del Sistema de Medición de la Calidad de la Enseñanza. UCMaule, *Revista Académica*, 54, 65-90. doi:10.29035/ucmaule.54.65

Gómez, P., Pérez, C., Parra, P., Ortiz, L., Matus, O., McColl, P., ... Meyer, A. (2015). Relación entre el bienestar y el rendimiento académico en alumnos de primer año de medicina. *Revista médica de Chile*, 143(7), 930-937. doi:10.4067/S0034-98872015000700015

González, A., & Vera, E. (2018). El nivel de ansiedad de los estudiantes extranjeros y el efecto en su rendimiento académico. *Revista Experiencia Docente*, 5(1), 84-102. Recuperado de <http://experienciadocente.ecci.edu.co/index.php/experienciadoc/article/view/80/EL%20NIVEL%20DE%20ANSIEDAD%20DE%20LOS%20ESTUDIANTE%20EXTRANJEROS%20Y%20EL%20EFECTO%20EN%20SU%20RENDIMIENTO%20ACAD%20MICO>

González, M.I., Álvarez, A.Y., & Nashiki, R.M. (2015). *Evaluación de los aprendizajes en la licenciatura en psicología educativa: diseño de una prueba objetiva para el curso de psicología evolutiva de la adultez y vejez*. Memoria electrónica del XIII Congreso Nacional de Investigación Educativa. Recuperado de <http://xplora.ajusco.upn.mx:8080/xplora-pdf/0340.pdf>

- Goñi, E., Ros, I., & Fernández-Lasarte, O. (2018). Academic performance and school engagement among secondary school students in accordance with place of birth, gender and age. *European Journal of Education and Psychology*, 11(2), 93-105. Recuperado de <https://dialnet.unirioja.es/servlet/articulo?codigo=6683415>
- Goode, W.J., & Hatt, P. K. (1998). *Métodos de investigación social*. México: Trillas.
- Gregory, R.J. (2001). Evaluación psicológica. Historia, principios y aplicaciones. México: El Manual Moderno.
- Hernández, R., Fernández, C., & Baptista, P. (2006). *Metodología de la Investigación*. México: Mc Graw Hill.
- Henrysson, S. (1971). Gathering, analysing, and using data on test items. En R.L. Thorndike (Ed.), *Educational Measurement* (pp. 130–159). Washington, DC: American Council on Education.
- Hincapié, D.A., Ramos, A., & Chrino-Barceló, V. (2018). Aprendizaje Basado en Problemas como estrategia de Aprendizaje Activo y su incidencia en el rendimiento académico y Pensamiento Crítico de estudiantes de Medicina. *Revista Complutense de Educación*, 29(3), 665-681.
doi:10.5209/RCED.53581
- Hogan, T.P. (2004). *Pruebas psicológicas*. México: El Manual Moderno
- Horna, A.P. (2018). *Influencia del estado nutricional en el rendimiento académico en escolares de 6 a 10 años en la escuela particular “Thomas Russell Crampton” de la ciudad de Cayambe, 2017*. (Tesis de grado inédita).

Universidad Técnica del Norte, Ecuador. Recuperada de
<http://repositorio.utn.edu.ec/bitstream/123456789/8107/1/06%20NUT%20243%20TRABAJO%20DE%20GRADO.pdf>

Huerta, A., & Pantoja, A. (2016). Efectos de un programa educativo basado en el uso de las TIC sobre el rendimiento académico y la motivación del alumnado en la asignatura de tecnología de educación secundaria. *Educación XX1*, 19(2), 229-250. doi:10.5944/educXX1.14224

Iniguez-Monroy, C.G., Aguilar-Salinas, W.E., De Las Fuentes-Lara, M., & Rodriguez-Gonzalez, R.E. (2017). El Interés en la Química General para Ingenierías y el Bajo Rendimiento Escolar. *Formación Universitaria*, 10(4), 33-42. doi:10.4067/S0718-50062017000400004

Juárez, C.S. (2015). *Monografía para el diseño, elaboración y análisis de los reactivos de un instrumento*. Centro Universitario UAEM Ecatepec.

Kerlinger, F.N. & Lee, H.B. (2001). *Investigación del Comportamiento. Métodos de investigación en ciencias sociales*. México: Mc GrawHill.

Korzeniowski, C., Cupani, M., Ison, M., & Difabio, H. (2016). Rendimiento escolar y condiciones de pobreza: el rol mediador de las funciones ejecutivas. *Electronic Journal of Research in Educational Psychology*, 14(3), 474-494. doi:10.14204/ejrep.40.15152

Lamas, H.A. (2015). Sobre el rendimiento escolar. *Propósitos y Representaciones*, 3(1), 313-386. doi:10.20511/pyr2015.v3n1.74

Magnusson, D. (2009). *Teoría de los Tests*. México: Trillas.

Manzano-Carrasco, S., López-Serrano, S., Suárez-Manzano, S., & Ruiz-Ariza, A.

(2018). Análisis descriptivo y de relación entre la actividad física y el rendimiento académico en jóvenes estudiantes del centro de España.

International Journal of Developmental and Educational Psychology, 1(2), 223-228. Recuperado de

<https://dialnet.unirioja.es/servlet/articulo?codigo=6432664>

Marín, M., Carballo, J.L., & Coloma-Carmona, A. (2018). Rendimiento académico y cognitivo en el uso problemático de Internet. *Adicciones*, 30(2), 101-110.

Recuperado de

<http://www.adicciones.es/index.php/adicciones/article/view/844/935>

Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice*, 14(4), 5-8. doi:10.1111/j.1745-3992.1995.tb00881.x

Morales, L.A., Morales, V., & Holguín, S. (2016). Rendimiento escolar. *Revista Electrónica de Humanidades, Tecnología y Ciencia*, 15, 1-5. Recuperado de <https://docplayer.es/57397498-Rendimiento-escolar-leticia-andrea-morales-sanchez-esiqie-ipn-virginia-morales-sanchez-upiicsa-ipn.html>

Moreno, R., Martínez, R.J., & Muñiz, J. (2004). Directrices para la construcción de ítems de elección múltiple. *Psicothema*, 16(3), 490-497. Recuperado de <http://www.psicothema.com/psicothema.asp?id=3023>

Muñiz, J. (2003). *Teoría Clásica de los Test*. Madrid, España: Ediciones Pirámide.

- Navarro, S.J., & Blandon, S.L. (2017). Determinantes que inciden en la calidad de rendimiento académico de los estudiantes de ingeniería. *Revista Científica de FAREM-Estelí. Medio ambiente, tecnología y desarrollo humano*, 6(24), 126-142. doi:10.5377/farem.v0i24.5556
- Pulido, F., & Herrera, F. (2018). Influencia de los estados emocionales contrapuestos en el rendimiento académico. *Docencia e Investigación*, XLIII(29), 45-66. Recuperado de <https://revista.uclm.es/index.php/rdi/article/view/1710/1459>
- Quintana, S.O., & Numa, J. (2018). hábitos de estudio, métodos de estudio y rendimiento académico en los estudiantes de primer semestre de tecnología en gestión comercial y financiera y administración de empresas, año 2014, de la Universidad Francisco de Paula Santander, seccional Ocaña desde la propuesta de Fernández Pozar. *Journal of Social Sciences and Management Research Review*, 1(1), 77-91. Recuperado de <http://scmjournals.com/ojs/index.php/jscmrr/article/view/9/29>
- Regueiro, B., Suárez, N., Estévez, I., Rodríguez, S., Piñeiro, I., & Valle, A. (2018). Deberes escolares y rendimiento académico: un estudio comparativo entre el alumnado inmigrante y nativo. *Revista de Psicología y Educación*, 13(2), 92-98. doi:10.23923/rpye2018.01.160
- Rico, A., & Sánchez, D. (2018). Diseño de un modelo para automatizar la predicción del rendimiento académico en estudiantes del IPN. *Revista*

Iberoamericana para la Investigación y el Desarrollo Educativo, 8(16), 1-21.

doi: 10.23913/ride.v8i16.340

Rivas, T., González, M.J. & Delgado, M. (2010). Descripción y Propiedades Psicométricas del Test de Evaluación del Rendimiento Académico (TERA).

Interamerican Journal of Psychology, 4(2), 279-290. Recuperado de

<http://www.redalyc.org/articulo.oa?id=28420641009>

Roux, E., & Anzures, E.E. (2015). Estrategias de aprendizaje y su relación con el rendimiento académico en estudiantes de una escuela privada de educación media superior. *Revista Electrónica "Actualidades Investigativas en Educación"*, 15(1), 1-16. doi:10.15517/aie.v15i1.17731

Salkind, N.J. (1999). *Métodos de Investigación*. México: Prentice Hall.

Sánchez, H., & Reyes, C. (2015). *Elaboración y validación de una prueba de comprensión numérica para alumnos del nivel primario (PCN-NEP) de instituciones educativas de Santiago de Surco*. (Tesis de maestría inédita). Universidad Ricardo Palma, Perú. Recuperada de <http://repositorio.urp.edu.pe/bitstream/handle/urp/837/SanchezReyes2015.pdf?sequence=1&isAllowed=y>

Sanz, I., Fernández-Martínez, I., Espada, J.P., & Orgilés, M. (2018). ¿Tienen los preadolescentes con exceso de peso y baja autoestima peor rendimiento académico? Un estudio transversal con una muestra española. *Revista de Psicología Clínica con Niños y Adolescentes*, 5(5), 54-58. Recuperado de http://www.revistapcna.com/sites/default/files/08_0.pdf

- Sarrazola-Moncada, A.M., Soto-Fadul, J.D., Carmona, L., García, M., Rojas, G., Tabares, V., & Vásquez, M. (2016). Trastornos emocionales y rendimiento académico en estudiantes de odontología. *Revista Colombiana de Investigación en Odontología*, 7(20), 106-116. doi:10.25063/21457735.264
- Serrano, A., Mérida, R., & Taberner, C. (2016). La autoestima infantil, la edad, el sexo y el nivel socioeconómico como predictores del rendimiento académico. *Revista de Investigación en Educación*, 14(1), 33-66.
- Recuperado de <http://reined.webs.uvigo.es/index.php/reined/article/view/224/245>
- Sifuentes, P.A., Pesantes, C.R., Rojas, J.L., Falcón, A.N., & Manes, G.A. (2017). Applets con geogebra en el rendimiento académico en matemática de los estudiantes del II ciclo escuela profesional de administración. *Big Bang Faustiniiano*, 6(4), 20-23. Recuperado de <http://revistas.unjpsc.edu.pe/index.php/BIGBANG/article/view/177/179>
- Sizalima-Cuenca, S.M., Vásquez-Mena, A.E., Maldonado-Vélez, B.T., & Ávila-Sánchez, D.E. (2017). La calidad de la clase como eje articular en el rendimiento académico de los estudiantes. *OLIMPIA. Revista de la Facultad de Cultura Física de la Universidad de Granma*, 14(45), 252-266.
- Recuperado de <https://dialnet.unirioja.es/servlet/articulo?codigo=6210521>
- Soliz, N., Mena, V., & Lara, T. (2017). El consumo de alcohol y el rendimiento académico de los estudiantes de la facultad de ciencias económicas de la Universidad Central del Ecuador en el año 2015. *Revista Publicando*, 4(10-

2), 120-142. Recuperado de

https://rmlconsultores.com/revista/index.php/crv/article/view/481/pdf_307

Tapia, V., & Luna, J. (2010). Validación de una prueba de habilidades de pensamiento para alumnos de cuarto y quinto de secundaria y primer año de universidad. *Revista de Investigación en Psicología*, 13(2), 17-59.

Recuperado de

<http://revistasinvestigacion.unmsm.edu.pe/index.php/psico/article/view/3710/2974>

Trelles, H.J., Alvarado, H.P., & Montánchez, M.L. (2018) Estrategias y estilos de aprendizaje y su relación con el rendimiento académico en estudiantes universitarios de Psicología Educativa. *Revista Killkana Sociales*, 2(2), 9-16. doi:10.26871/killkana_social.v2i2.292

Valle, A., Pan, I., Núñez, J.C., Rosário, P., Rodríguez, S., & Regueiro, B. (2015). Deberes escolares y rendimiento académico en Educación Primaria. *Anales de Psicología*, 31(2), 562-569. doi:10.6018/analesps.31.2.171131

- **El presente libro está organizado en dos capítulos. El capítulo uno denominado “Rendimiento académico” y el capítulo dos intitulado “Validación de pruebas de rendimiento académico”.**
- **En el primero se expone el marco prescriptivo y conceptual que circunscribe y le otorga un sentido referencial al contenido central del presente libro.**
- **En el capítulo dos se abordan la Confiabilidad, la Validez de contenido y el Análisis de reactivos como análisis centrales, valga la redundancia, que permiten determinar las propiedades psicométrica de una prueba de rendimiento académico.**