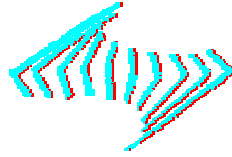


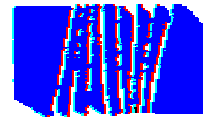


Universidad Centroccidental  
"Lisandro Alvarado"



Universidad Nacional Experimental  
Politécnica Antonio José de Sucre

U  
N  
E  
X  
P  
O



Universidad Pedagógica  
Experimental Libertador

## Programa Interinstitucional Doctorado en Educación

Carlos Ruiz Bolívar, PhD  
Profesor Titular UPEL / PIDE  
[cruizbol@intercable.net.ve](mailto:cruizbol@intercable.net.ve)

### VALIDEZ

En la sección anterior, nuestro interés se centró en determinar hasta dónde los resultados de un instrumento de medición son estables a través del tiempo, cuando éste ha sido utilizado varias veces con los mismos sujetos y bajo las mismas condiciones de aplicación. En esta sección, nos interesa estudiar la exactitud con que pueden hacerse mediciones significativas y adecuadas con un instrumento, en el sentido de que mida realmente el rasgo que pretende medir. Esta propiedad o característica de un instrumento de medición recibe el nombre de *validez*. Es decir, en sentido general, la validez de un instrumento tiene que ver con las preguntas siguientes: ¿qué miden los puntajes del test? y ¿qué predicen dichas puntuaciones? (Guilford, 1954; Nunnally, 1967; Anastasi, 1976; Magnusson, 1982).

Por ejemplo, si un rasgo, como la inteligencia académica, es significativo o relevante para el éxito en el aprendizaje formal, construimos un instrumento para medir dicho rasgo y lo usamos en las tareas de diagnóstico, selección y de orientación escolar o académica. Pero, obviamente, es necesario que el instrumento mida el rasgo que hemos encontrado relevante en la situación donde el mismo va ser utilizado y no alguna otra característica.

La validez de un instrumento, por lo general, no constituye un problema en el caso de la medida de los objetos físicos, tales como longitud, peso, capacidad. Por supuesto que la estatura de una persona se mide con una cinta métrica y el peso de un objeto, con una balanza (siempre que la balanza funcione adecuadamente). Sin embargo, con los métodos usados para medir variables psicoeducativas, es necesario probar empíricamente que el instrumento es válido en todos los casos.

Cuando elaboramos una escala para medir la actitud de los docentes hacia la innovación educativa, debemos probar que los puntajes de la escala realmente distinguen entre aquellos docentes que tienen una actitud favorable hacia la

innovación y aquellos cuya actitud es desfavorable. Algunas veces, puede parecer obvio que un instrumento mida un determinado rasgo; sin embargo, pudiera ocurrir que cuando se le someta a una evaluación empírica, el instrumento esté midiendo algún rasgo distinto. De allí que sea necesario tener presente que la validez no es materia de presunción sino de demostración empírica.

Cuando estimamos la validez de un instrumento, necesitamos saber qué característica deseamos que prediga. Este rasgo se llama *variable criterio*. Nos interesa saber qué tan bien corresponden las posiciones de los individuos en las distribuciones de los puntajes obtenidos con respecto a sus posiciones en el continuo que representa la variable criterio. Por lo general, la validez es estimada a través de una correlación entre los puntajes de una medida y las puntuaciones de la variable criterio, la cual predice. Este índice se denomina coeficiente de validez.

Un instrumento no tiene un coeficiente fijo de validez que sirva para cualquier propósito y para cualquier grupo de individuos. La validez de un instrumento varía de acuerdo con el propósito con que se use y el grupo dentro del cual discrimina. Por ejemplo, si una prueba de aptitud académica predice el rendimiento (variable criterio) de los estudiantes en el primer semestre de la universidad, es de esperarse que la mayoría de los estudiantes que obtuvieron altas calificaciones en la prueba de aptitud, también sean los que obtengan más éxito en dicho lapso académico. Sin embargo, el hecho de que la prueba de aptitud académica haga una predicción razonable del rendimiento universitario, no garantiza que también pueda predecir la motivación al logro de dichos estudiantes, o quiénes se graduarán al final de la carrera, ni muchos menos quiénes tendrán éxito en sus respectivos campos profesionales.

### **Tipos de Validez**

La validez como la confiabilidad de un instrumento, a pesar de tener ambas sus respectivas definiciones teóricas genéricas, en la práctica, dado que estas características pueden adoptar diferentes significados, es necesario adjetivarlas cada vez que se haga referencia a ellas. Es decir, no es suficiente con que se diga que un instrumento determinado tiene una alta confiabilidad o validez. Por qué el lector especializado inmediatamente se preguntaría ¿qué tipo de confiabilidad? o ¿qué tipo de validez? Es necesario ser específico al usar estos términos. De esta manera se podría decir, por ejemplo, este instrumento tiene una alta confiabilidad de consistencia interna o esta prueba tiene una alta validez predictiva.

En el caso particular de la validez, cuando investigamos esta característica en un instrumento determinado, intentamos responder tres tipos de cuestiones, que aluden a igual número de tipos de validez. Estas cuestiones son:

1. ¿Cuán representativo es el comportamiento elegido como muestra del universo que se intenta representar?
2. ¿Qué significado tiene el comportamiento con respecto a los atributos del individuo que son de interés para la medición?

3. ¿Hasta dónde se puede predecir el rendimiento del sujeto o su aprendizaje en un programa de entrenamiento (o hasta dónde se puede anticipar su nivel de desempeño en el trabajo), a partir de su ejecución en la prueba?

Estos tres tipos de preguntas corresponden a lo que en la literatura especializada se ha denominado respectivamente como *validez de contenido*, *validez de constructo* y *validez predictiva*. A continuación se presenta una descripción de cada una de ellas, así como sus respectivos procedimientos para estimar los diferentes índices de validez.

### **Validez de Contenido**

A través de la validez de contenido se trata de determinar hasta dónde los ítems de un instrumento son representativos del dominio o universo de contenido de la propiedad que se desea medir. Esta definición, aunque bien en teoría, presenta dificultades prácticas, ya que es imposible extraer muestras aleatorias de reactivos de un universo de contenido, puesto que este existe sólo conceptual o teóricamente.

Algunos autores han tratado de resolver el problema, de la validez de contenido, generando grandes cantidades de ítems de un dominio determinado, para luego obtener muestras representativas de dicho universo, a los fines de integrar un instrumento; sin embargo, en la opinión de autores como Kerlinger (2002), la validez de contenido de tales conjuntos, independientemente de qué tan grandes y tan “buenos” sean los reactivos, es siempre dudosa y, por tanto, cuestionable. ¿Cómo lograr, entonces, una estimación de la validez de contenido de un instrumento?

Ítem	Congruencia		Claridad		Tendenciosidad		Observaciones
	Si	No	Si	No	Si	No	
<b>1</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____
<b>2</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____
<b>3</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____
<b>4</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____
<b>5</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____
<b>6</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____
<b>7</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____
<b>8</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____
<b>9</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____
<b>10</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____
<b>11</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____

Figura 1. Planilla de Validación

A diferencia de otros tipos de validez, la de contenido no puede ser expresada cuantitativamente, a través de un índice o coeficiente; ella es más bien una cuestión de juicio. Es decir, la validez de contenido, por lo general, se estima de manera subjetiva o intersubjetiva. El procedimiento más comúnmente empleado para determinar este tipo de validez, es el que se conoce con el nombre de **juicios de expertos**, para lo cual se procede de la siguiente manera:

1. Se seleccionan dos jueces o expertos, por lo menos, a los fines de juzgar, de manera independiente, la “bondad” de los ítemes del instrumento, en términos de la relevancia o congruencia de los reactivos con el universo de contenido,

la claridad en la redacción y la tendenciosidad o sesgo en la formulación de los ítemes.

2. Cada experto recibe suficiente información escrita acerca de: (a) el propósito de la prueba; (b) conceptualización del universo de contenido; (c) plan de operacionalización o tabla de especificaciones (en el caso de las pruebas de rendimiento académico).
3. Cada juez recibe un instrumento de validación en el cual se recoge la información de cada experto. Dicho instrumento normalmente contiene las siguientes categorías de información por cada ítem: congruencia ítem-dominio, claridad, tendenciosidad y observaciones (ver Figura 2).
4. Se recogen y analizan los instrumentos de validación y se toman las decisiones siguientes: (a) los ítemes donde hay un 100 por ciento de coincidencia favorable entre los jueces (los ítemes son congruentes, están escritos claramente y no son tendenciosos) quedan incluido en el instrumento; (b) los ítemes donde hay un 100 por ciento de coincidencia desfavorable entre los jueces, quedan excluidos del instrumento; y (c) los ítemes donde sólo hay coincidencia parcial entre los jueces deben ser revisados, reformulados, si es necesario, y nuevamente validados.

### **Validez de Constructo**

La validez de constructo intenta responder la pregunta ¿hasta dónde un instrumento mide realmente un determinado rasgo latente o una característica de las personas y con cuánta eficiencia lo hace? Esta pregunta tiene sentido, particularmente en los instrumentos que se utilizan en la investigación psicoeducativa, ya que, como se ha mencionado anteriormente, en este campo hacemos mediciones indirectas de ciertas variables internas del individuo que denominamos constructos. En consecuencia, es necesario que podamos mostrar evidencia de que, efectivamente, el instrumento mide el rasgo o constructo que pretende medir.

Gronlund (1976) señala que la validez de constructo interesa cuando queremos utilizar el desempeño de los sujetos con el instrumento para inferir la posesión de ciertos rasgos o cualidades psicológicas. Por ejemplo, en lugar de hablar de los puntajes que una persona obtuvo en un instrumento determinado, podemos hablar de: (a) la actitud hacia la matemática; (b) la satisfacción o disfrute con el aprendizaje de la matemática; y (c) la valoración o significado de la matemática. Todas éstas son cualidades hipotéticas llamadas construcciones, cuya existencia se supone para explicar la conducta en muchas y diferentes situaciones específicas.

1. La relación entre la conceptualización teórica del instrumento y su estructura factorial. Por ejemplo, en el desarrollo de la escala AC-2000, que mide el Auto-concepto de los alumnos de sexto grado, Ruiz (1988) comprobó, a través de la técnica del **Análisis Factorial**, que las tres dimensiones hipotetizadas en la

escala (Autoconcepto Social, Autoconcepto Personal y Autoconcepto Escolar) tenían soporte empírico en los datos (ver Cuadro 1).

Para estudiar la validez de constructo de un instrumento es necesario que exista una conceptualización clara del rasgo bajo estudio, con base en una “teoría” determinada. Esta nos permitirá tener una idea clara acerca de cómo se manifiesta el atributo bajo estudio, qué tipo de rendimiento, en la vida diaria, se facilitan por la posesión del atributo, cuáles subgrupos de la población lo poseen en alto o bajo grado y qué condiciones favorecen o impiden su expresión. La teoría del atributo, también sugiere las tareas de prueba que son las apropiadas para hacer aflorar el atributo. Además, la teoría sugiere las clases de evidencias que deben considerarse para evaluar las bondades de la prueba bajo estudio, si es que de hecho hace aflorar o depende directamente del constructo que se analiza.

Cronbach (1960) ha sugerido los pasos siguientes para establecer la validez de constructo: (a) identificar las construcciones que pudieran explicar la ejecución en el instrumento; (b) formulación de hipótesis comprobables a partir de la teoría que enmarca a cada construcción; y (c) recopilación de datos para probar estas hipótesis. Estas hipótesis se pueden enunciar en relación con cualquiera de los siguientes tipos generales de evidencia (Thorndike, 1989; Helmstadter, 1964):

Cuadro 1

Estructura factorial de la Escala AC-2000

Item	Factor 1 A. Social	Item	Factor 2 A. Escolar	Item	Factor 3 A. Personal
<b>3</b>	0,51	<b>4</b>	0,52	<b>1</b>	0,43
<b>6</b>	0,63	<b>8</b>	0,43	<b>2</b>	0,48
<b>16</b>	0,49	<b>12</b>	0,45	<b>5</b>	0,58
<b>22</b>	0,48	<b>25</b>	0,54	<b>7</b>	0,43
<b>26</b>	0,64	<b>28</b>	0,57	<b>9</b>	0,47
<b>27</b>	0,54	<b>29</b>	0,55	<b>10</b>	0,56
		<b>30</b>	0,40	<b>13</b>	0,41

2. La información correlacional, esto es, que muestre las correlaciones entre la prueba bajo estudio y otras medidas (pruebas o eventos en la vida real) que reflejen o dependan del atributo en cuestión. Por ejemplo, en el mismo estudio de Ruiz (1988) se encontró una correlación de 0,79 entre el instrumento AC-2000 y la escala de Auto-estima de Coopersmith (1959).

También se pueden formular y evaluar hipótesis que planteen relaciones con otras variables en las que teóricamente: (a) es esperable una correlación positiva con el constructo bajo estudio, conocido en la literatura con el nombre de *validez convergente*; y (b) no es esperable ningún tipo de relación con el constructo objeto de validación, conocida como *validez discriminante* (Campbell y Fiske, 1959).

3. Los datos sobre las diferencias entre grupos, comparando las puntuaciones de prueba de los subgrupos que se puede esperar difieren en el nivel del atributo. Por ejemplo, en el desarrollo de una escala de Auto-concepto, se podría probar la hipótesis de que los estudiantes con alto y bajo auto-concepto académico difieren en rendimiento escolar.
4. La información que muestre los efectos de tratamiento o intervenciones experimentales que se puede esperar influyan en la expresión del atributo. Por ejemplo, existen evidencias en la literatura que brindan soporte a la hipótesis de que una estrategia instruccional centrada en la retroalimentación positiva y crítica en las pruebas de evaluación formativa, de estudiantes de sexto grado, mejora el Auto-concepto de dichos alumnos (ver Sánchez de Hurtado, 1994).
5. Consistencia interna. En este sentido, podríamos predecir correlaciones altas entre ítemes debido a que todos ellos supuestamente miden el mismo constructo: Auto-concepto.

### **Validez Predictiva**

La palabra predicción o predictivo normalmente se la asocia con visión o anticipación de futuro. En este sentido, cuando estudiamos la validez predictiva de un instrumento lo que nos interesa es determinar hasta dónde podemos anticipar el desempeño futuro de una persona en una actividad determinada, a partir de su ejecución actual en dicho instrumento; por ejemplo, se podría estudiar, hasta dónde la Prueba de Aptitud del Consejo Nacional de Universidades predice el éxito académico de los estudiantes en los primeros semestres universitarios, o anticipar el desempeño futuro de un vendedor a partir de su ejecución en un test de inteligencia social.

En consecuencia, la validez predictiva, también llamada validez de criterio externo o validez empírica, se estudia comparando los puntajes de un instrumento (variable independiente) con una o más variables externas (variables dependientes) denominadas variables criterio. Se asume que tales criterios, indicadores del desempeño futuro, están teóricamente y lógicamente relacionados con el rasgo representado en el instrumento bajo estudio. Esta comparación entre los puntajes de la variable en estudio y los de la variable criterio se expresa a través de un coeficiente de correlación, el cual se interpreta como un índice de validez. Entre más alta sea la correlación entre una medida o medidas de aptitud académica y el promedio de notas, tomado como variable criterio, mejor será la validez predictiva de la prueba de aptitud académica.

Es importante destacar que la validez predictiva suele estar asociada con problemas y resultados prácticos; es decir, el interés no es tanto en lo que está detrás del desempeño en la prueba, sino más bien en ayudar a resolver problemas prácticos y tomar decisiones. Muchos de estos problemas y toma de decisiones están relacionados con la evaluación, selección y asignación de personas para diferentes actividades (estudio, trabajo, deporte, arte, etc).



No obstante, determinar la validez predictiva de un instrumento tiene sus problemas, los cuales en buena medida están relacionados con la calidad del criterio externo seleccionado, ya que si éste no es significativo o relevante con respecto a lo que mide el instrumento bajo estudio, la validez predictiva no podrá quedar evidenciada, pero no porque efectivamente la prueba no tenga un determinado tipo de validez predictiva, sino porque el criterio utilizado, para probar dicha propiedad del instrumento, no fue el apropiado.

Es decir, la obtención de criterios apropiados para validar una prueba a veces tiene sus dificultades; por ejemplo, ¿cuál de los siguientes podría ser el mejor criterio para predecir el nivel de eficiencia de un docente? ¿el nivel de dominio que los alumnos tienen de los objetivos instruccionales? ¿el promedio de calificaciones? ¿la tasa de estudiantes aprobados o promovidos? ¿la claridad con que el docente expone sus clases?, ¿el dominio que el docente tiene del contenido de la asignatura que enseña? ¿la responsabilidad del docente en el cumplimiento de las tareas inherentes a su cargo?

Como se pudo observar en el párrafo anterior, todos los criterios mencionados pudieran estar, de alguna manera, relacionados con la eficiencia de un docente; el problema está en cómo determinar el más relevante, o los más relevantes, porque puede darse el caso de que el desempeño futuro que se aspira predecir, a partir de la ejecución actual en una prueba, se exprese no a través de un criterio sino de varios, como pareciera ser el caso, ya mencionado, de la eficiencia de un docente.

Lo anterior indica que, en algunos casos, existen criterios múltiples, como también pueden existir los predictores múltiples. De hecho, dada la complejidad multifactorial que determina el comportamiento humano, en cualquier actividad donde se desempeñe, la tendencia actual, en los procedimientos de selección (de estudiantes, trabajadores, deportistas, artistas, etc) es a utilizar predictores múltiples (v.g. baterías de pruebas), más que instrumentos únicos. De la misma manera, y por la misma razón, cada vez más se utilizan criterios múltiples como indicadores del desempeño, más que criterios simples o únicos.

### **Factores que Afectan la Validez de una Prueba**

La validez de los resultados de un instrumento de medición puede verse afectada por diferentes factores, los cuales pueden estar asociados tanto al proceso de elaboración de la prueba misma, como a otros factores relacionados con los procedimientos de administración, calificación e interpretación de los resultados. A continuación se presentan algunos de estos factores con la intención deliberada de que, en la medida de lo posible, sean evitados en el proceso de construcción de los instrumentos. Entre estos factores están:

1. *Construcción del instrumento.* Estos factores pueden influir en que las preguntas del instrumento no funcionen en la forma prevista y, por lo tanto, contribuyen a minar la validez de los resultados; algunos de ellos son: (a)

instrucciones imprecisas o vagas; (b) estructura sintáctica de la oración demasiado difícil; (c) preguntas con niveles de dificultad inapropiados; (d) preguntas que sugieren la respuesta; (e) ambigüedad en la formulación de los reactivos, lo cual lleva a diferentes interpretaciones de los mismos; (f) pruebas demasiado cortas; (g) ítemes incongruentes con el dominio o universo de contenido; (h) ordenamiento inadecuado de los ítemes; (i) patrón identificable de respuestas, particularmente en las preguntas de selección.

2. *Administración y calificación de la prueba.* Entre estos factores se encuentran: tiempo insuficiente para responder, ayuda adicional a algunos sujetos, más allá de las instrucciones generales y de los ejercicios de práctica, uso de “chuletas” o ayudas escritas, por parte de los sujetos, uso de la subjetividad de la puntuación de las preguntas.
3. *Respuestas de los sujetos.* Algunos de estos factores son: bloqueo de los sujetos para responder, debido a situaciones emocionales y las respuestas formuladas al azar.
4. *Naturaleza del grupo y del criterio.* Como ya se ha mencionado, la validez es siempre específica con respecto a un grupo en particular. Por lo tanto, si una prueba de comprensión de lectura, por ejemplo, que ha sido diseñada y validada para alumnos de sexto grado, pudiera carecer de validez de contenido y predictiva, para medir el nivel de comprensión de lectura en estudiantes de noveno grado.

Por otra parte, es sumamente importante prestar atención al criterio que va a ser seleccionado como validador del instrumento, ya que, como se mencionó anteriormente, el uso de un criterio inapropiado puede disfrazar o distorsionar la validez de la prueba.

### **La Validez como Varianza Común**

En la sección anterior se señaló que la confiabilidad, genéricamente hablando, podía ser interpretada como la correlación del instrumento consigo mismo, y que, además, podía ser definida como la proporción de la varianza verdadera ( $V_v$ ) sobre la varianza total; o sea como:

$$r_{tt} = \frac{V_v}{V_t}$$

De la misma manera, la validez puede ser interpretada como la relación que un instrumento, bajo estudio, tiene con otro que se utiliza como criterio y puede ser definida como el porcentaje de varianza común que dicho instrumento comparte con el criterio, como se expresa en la fórmula siguiente:

$$Val = \frac{V_{co}}{V_t}$$

En donde:

$V_{al}$  = validez del instrumento;

$V_{co}$  =varianza común que el instrumento comparte con el criterio; y

$V_t$  = varianza total del instrumento bajo estudio.

Pero, ¿Qué se entiende por varianza común? ¿Cómo se relaciona la varianza verdadera ( $V_v$ ) con la varianza común ( $V_{co}$ )? ¿Qué se entiende por varianza específica de una prueba? ¿Qué es la varianza de error? Para intentar responder estas interrogantes presentaremos algunas nociones básicas e introductorias de la teoría factorial.

De acuerdo con esta teoría, cuando dos pruebas están relacionadas, como ocurre en el caso de la validación de un instrumento, pueden ser identificados varios tipos de varianzas, como son: (a) la *varianza de factor común*; es decir, el porcentaje de varianza que un instrumento comparte con otro; (b) la *varianza específica*, que es la varianza sistemática de una medida que no es compartida con otra; y (c) la *varianza de error*, que es una varianza aleatoria (no sistemática) debido a factores inherentes a la construcción de la prueba, a sus condiciones de aplicación o a situaciones personales presentadas por el sujeto en el momento de responder el instrumento.

Es decir, si una prueba mide habilidades que también miden otras pruebas, se tiene una varianza de factor común; por el contrario, si la prueba mide una habilidad que no mide ninguna otra prueba, se tiene una varianza específica (Kerlinger, 1988).

De acuerdo con esto, la varianza total de una medida está integrada por varios componentes, a saber: varianza de factor común, varianza específica y varianza de error, lo cual puede ser expresado mediante la siguiente ecuación:

$$V_t = V_{co} + V_{es} + V_e$$

Esta relación puede ser representada también gráficamente, como se indica a través de la Figura 2.

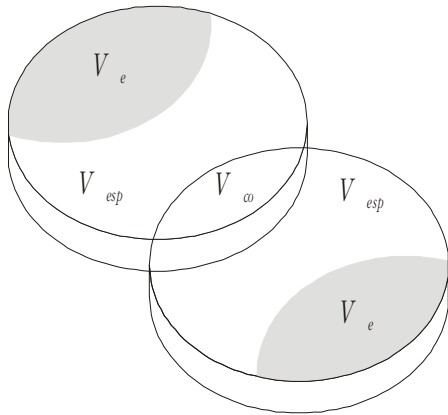


Figura 2. Componentes de la variancia de dos instrumentos (Adaptado de Kerlinger, 2002)

En cuanto a la relación entre la variancia verdadera ( $V_v$ ) y la variancia común, observemos que la variancia total, por una parte, puede ser expresada como la sumatoria de la variancia verdadera, más la variancia de error, como se presenta en la siguiente ecuación:

$$V_t = V_v + V_e$$

Pero, al mismo tiempo, la variancia total puede ser expresada como la sumatoria de la variancia de factor común, más la variancia específica, más la variancia de error, como se señala en la siguiente ecuación:

$$V_t = V_{co} + V_{es} + V_e$$

De lo anterior se deduce que la variancia verdadera ( $V_v$ ) está integrada por la variancia común ( $V_{co}$ ), más la variancia específica ( $V_{es}$ ), como se presenta en la Figura 3.

$$\text{Variancia Total} = \underbrace{\text{Varianza Verdadera}}_{\text{Varianza Común} + \text{Varianza Específica}} + \text{Varianza de Error}$$

Figura 3. Componentes de la Varianza

### Relación entre Validez y Confiabilidad

Si la confiabilidad puede ser interpretada como la proporción de la variancia verdadera ( $V_v$ ) con respecto a la variancia total del instrumento, como se expresa en la ecuación:

$$r_{tt} = \frac{V_v}{V_t}$$

Y, por otra parte, la validez es definida como la proporción de la varianza común con respecto a la varianza total, como se indica en la ecuación:

$$V_{al} = \frac{V_{co}}{V_t}$$

Si sabemos, además, que la varianza total ( $V_t$ ) puede ser descompuesta en sus componentes, como se expresa en la fórmula:

$$\frac{V_t}{V_t} = \frac{V_v}{V_t} + \frac{V_e}{V_t}$$

Pero, a su vez, la varianza verdadera ( $V_v$ ) está compuesta por la varianza común, más la varianza específica ( $V_{es}$ ), la ecuación anterior puede ser reformulada como sigue:

$$\frac{V_t}{V_t} = \frac{V_{co}}{V_t} + \frac{V_{es}}{V_t} + \frac{V_e}{V_t}$$

De lo anterior se deduce que la proporción de varianza que comprende la interpretación de la confiabilidad  $r_{tt} = (V_{co} + V_{es})/V_t$ , es mayor que la proporción de varianza incluida en la interpretación de la validez  $V_{al} = V_{co}/V_{es}$ . No obstante, desde el punto de vista teórico, pudiera darse la circunstancia en que un instrumento sólo tuviera  $V_{co} + V_e$  y no tuviera  $V_{es}$ , en este caso la proporción de varianza que define la confiabilidad sería igual a la que define a la validez. Por lo tanto, en términos de magnitudes es de esperarse que, bajo condiciones normales, el coeficiente de confiabilidad de un instrumento siempre sea mayor o igual que su coeficiente de validez.

Ilustremos las relaciones anteriores con un ejemplo numérico. Supongamos que hemos diseñado una prueba para medir Vocabulario (PVO) en sexto grado y que hemos seleccionado una prueba de Comprensión de Lectura (PCL) como criterio externo de validación y encontramos que la varianza común entre PVO y PCL es igual a 0,35; mientras que la que comparte PCL con PVO es de 0,30.

Supongamos, además, que la varianza específica de PVO es de 0,20 y que la varianza de error es de 0,15. En tal situación se cumplen que: (a) las dos pruebas tienen una varianza compartida (Factor común A + B) de 0,65 (coeficiente de validez de PVO) ; (b) la varianza verdadera ( $V_{co} + V_{es}$ ) es de 0,85 (coeficiente de confiabilidad de PVO). Se cumple la tendencia de que en la mayoría de los casos la proporción de varianza que comprende la confiabilidad del instrumento es mayor que la que está implícita en la validez (Ver Cuadro 2)

## Cuadro 2

Relación entre los conceptos de confiabilidad y validez de acuerdo con los supuestos de la teoría factorial

Componentes de la Varianza	Vocabulario (A)	Comprensión de Lectura (B)
Factor Común	.35	.30
Varianza Compartida (Vco)	(A+B)	.65
Varianza Específica (Ves)	.20	
Varianza Verdadera (Vco+Ves)		.85
Varianza de Error (Ve)	.15	

### Importancia de la Validez

Una de las principales dificultades que se presenta con la validez de los instrumentos en Educación y Psicología, como ya se ha mencionado antes, es que la medición siempre es indirecta; es decir, debemos inferirla a partir del desempeño que los sujetos tienen en las pruebas. Este hecho en sí mismo representa un problema, ya que ¿cómo sabemos que estamos haciendo la inferencia correcta a partir de los resultados?. Pero, el problema de fondo sigue siendo el de poder asegurarse de que el instrumento mida lo que pretende medir, es decir, que sea válido a los fines propuestos. Muchos de los cuestionamientos que se hacen, hoy día, en la investigación sociológica y psicoeducativa tiene que ver, primero que todo, con la validez de la medida. Obviamente, un instrumento que no tiene una validez demostrada no vale la pena que sea utilizado.

Si comparamos la confiabilidad con la validez, nos damos cuenta que la obtención de la primera puede ser reducida básicamente a una cuestión técnica. Sin embargo, la validez es mucho más que eso. Tiene que ver con el aspecto sustantivo de la ciencia misma. También se relaciona con la epistemología, en tanto que teoría del conocimiento, y con los paradigmas científicos. Particularmente, en lo que se refiere a la validez de constructo, puesto que tiene que ver con la naturaleza de la “realidad” a las propiedades del objeto que se están midiendo (Kerlinger, 2002).

No obstante, las dificultades prácticas que se presentan para lograr obtener medidas válidas y confiables, dentro del paradigma de la ciencia clásica, en los últimos años se han desarrollado una serie de métodos, técnicas y procedimientos, que facilitan, cada vez más, esta tarea. Pero, más que el manejo de todo este instrumental tecnológico (métodos estadísticos, procedimientos electrónicos, paquetes computarizados, etc), lo más importante es que el investigador se haga consciente de la necesidad de utilizar instrumentos apropiados, técnicamente bien calibrados, a fin de garantizar la utilidad y significado de los resultados obtenidos.

Queda claro entonces que la construcción de instrumentos de medición no se reduce a la simple presentación de un listado de preguntas en un formato determinado. Construir “buenos” instrumentos de medición es, primero que todo, una tarea técnica, que requiere, por parte del investigador, un entrenamiento

específico para acometerla con éxito. Construir un instrumento técnicamente bien hecho implica, en sí mismo, una investigación. De allí que cuando se requiera hacer un estudio (trabajo o tesis de grado, trabajo de ascenso, investigación libre) antes de tomar la seria decisión de construir una instrumento de medición, sin ser un especialista en el área, se debería averiguar previamente acerca de la existencia de la disponibilidad comercial de dicho instrumento en el mercado, o a través de otros investigadores. Si después de esta indagación se llega a determinar que el instrumento no existe y que es indispensable trabajar en el desarrollo del mismo, lo más recomendable sería buscar el asesoramiento técnico especializado correspondiente.

## Referencias

- Anastasi, A. (1976). ***Psychological testing***. New York: MacMillan Publishing Co.
- Campbell, D. T., y Fiske, D. W. (1959). Convergent and discrimination validation by the multitrait multimethod matrix. ***Psychological Bulletin***, **56**, 81-105.ç
- Coopersmith, S. (1959). A method for determining self-esteem. ***Journal of abnormal and social psychology***, **59**, 87-94.
- Cronbach, L. J. (1960). ***Essentials or psychological testing***. New York: Harper and Row.
- Guilford, J. P. (1954). *Psychometric Methods*. New York: McGraw- Hill Book Co.
- Gronlund, N. E. (1976). ***Elaboración de test de aprovechamiento***. México: Trillas.
- Helmstadter, G. C. (1964). ***Principles of psychological measurement***. New York: Appleton.
- Kerlinger, F. N. (2002). ***Investigación del Comportamiento*** (3ra ed.). México: McGraw-Hill.
- Magnusson, D. (1982). ***Teoría de los tests***. México: Trillas.
- Nunnally, J. C. (1967). ***Psychometric theory***. New York: McGraw- Hill.
- Ruiz, B. C. (1988). Uso y Abuso de la escala Likert en la investigación psicoeducativa. ***Investigación y Posgrado***, **3**, **1**, 93-112.
- Thorndike, R. L. (1989). ***Psicometría aplicada***. México: Limusa.