

LA VALIDACION DE CONSTRUCTO: SU APLICACION AL CEED (CUESTIONARIO PARA LA EVALUACION DE LA EFICACIA DOCENTE)

JAVIER TOURÓN

INTRODUCCIÓN

El objetivo del presente trabajo es doble. De una parte, pretendemos realizar una serie de consideraciones conceptuales sobre la importancia de la validación de constructo en el ámbito pedagógico, con el ánimo de estimular la reflexión de los estudiosos e investigadores sobre un aspecto de gran importancia en la metodología de la investigación que, lamentablemente, cuidamos poco. De otra, presentar los resultados de la aplicación del análisis factorial clásico y de la matriz multirasgo-multimétodo a un instrumento diseñado para la evaluación de la eficacia validado en nuestro país, pero no publicado hasta ahora en castellano.

La metodología científico-experimental común, en buena medida, a campos de investigación cuyos objetos de estudio son muy diversos adquiere u la fisonomía y particularidades propias cuando se aplica al ámbito de la educación. Y ello es debido, principalmente, a los hechos que constituyen n nuestro centro de interés. Fenómenos de extraordinaria complejidad que, rara vez, pueden ser abordados con una seguridad razonable a través de indicadores simples, de variables aisladas.

Existirá un problema permanente con la medición de y en lo educativo: el hiato indudable entre la medida y lo que pretendemos medir. Si medir, al menos dentro de la teoría clásica, «es asignar números a objetos fenómenos o relaciones conforme a reglas » (cfr. Stevens, 1951), lo que habremos de procurar es reducir la distancia que de ordinario media entre el número y lo que suponemos que éste representa, y ello se facilita solamente cuando la medida se integra en la teoría.

La posibilidad de la aplicación de la medida en educación sufre las mismas objeciones que en el ámbito de la Psicología. Objeciones que Thorndike en la introducción a la edición de 1913 de su obra *An Introduction Theory of Mental and Social Measurements*, resume en tres apartados: 1) ausencia o imperfección de unidades de medida; 2) la falta de constancia en los hechos medidos, y 3) la extrema complejidad de los fenómenos medidos,, aunque, por fortuna, no son del todo insuperables (cfr. Thorndike, 191 , Jones, 1971; Yela, 1965; Pérez Juste, 1983; Orden Hoz, 1985), es conveniente tenerlas en mente como marco general de referencia cuando se pretende llevar a cabo un proceso de medición en nuestro ámbito.

Sabemos que se pueden distinguir, al menos, tres niveles diferenciada de medida, pero no excluyentes entre sí (cfr. López Peal, 1988). El primen, corresponde al proceso de asignación de números a las observaciones empíricas de acuerdo con unas determinadas reglas. Es el nivel previo para los dos siguientes (cfr. Guttman, 1971), con él se asocia la investigación univariable, siendo el correspondiente a la teoría clásica de la medición.

El segundo nivel corresponde a las medidas derivadas de las relaciones entre los valores de dos o más propiedades, establecidas a través de una hipótesis estructural, referida al universo de observables. Se asocia a la investigación correlacional, ya que en él tendremos medidas derivadas de las relaciones entre dos escalas que representan propiedades o atributos de un

mismo dominio o dominios diferentes. La medición es bivariante, de modo que las medidas son indicadores del grado y tipo de relación entre los valores de las variables que se relacionan (el enfoque multivariado podría considerarse una extensión del nivel bivariado).

Por su parte, el tercer nivel se refiere a las medidas de relación entre los datos empíricos, o las relaciones entre ellos, y un constructo subyacente a los mismos, representado a través de un modelo matemático. Es el nivel de medida teórico, que supone una nueva concepción del problema.

En efecto, se trata de establecer, desde la teoría, un modelo que refleje lo que el investigador entiende que es el constructo que desea medir, lo que se lleva a cabo con el planteamiento de una estructura matemática que postula determinadas relaciones entre las variables empíricas —que podrán medirse directamente— y las variables latentes con las que aquéllas se relacionan. El objetivo será establecer el grado de ajuste entre el modelo teórico propuesto y los datos empíricos obtenidos. La interdependencia con los dos niveles anteriores parece clara también.

Habitualmente el proceso estadístico para la comprobación de la bondad de ajuste entre los datos y el modelo propuesto puede realizarse a través, por ejemplo, de la aplicación de los modelos causales asociados a estructuras de covarianza como el LISREL (cfr., p. ej., Jóreskog, 1974, 1977; Jóreskog y Sörbom, 1976, 1983), que aquí no vamos a tratar.

Así como la teoría clásica se encuentra asociada a las escalas de medida tipo Stevens y, por ello, algunos autores como Krantz y Tversky (1971) la denominan estrategia numérica, el aspecto central de la teoría axiomática viene representado por el papel sustantivo que en la misma ocupa la teoría,

La medición es considerada una parte integrante de la teoría en lugar de un problema a ser superado antes de la construcción de la misma, de modo que la exactitud de la medida dependerá de la exactitud de la teoría. El enfoque axiomático está dirigido a probar la teoría propuesta, para lo cual intenta una medición conjunta de todas las variables implicadas.

El establecimiento de modelos juega aquí, como se señalaba, un papel primordial, de modo que se tratará de establecer una correspondencia entre el sistema relacional empírico (el mundo real) y un sistema relacional formal (el modelo). Este enfoque tiene una gran importancia, como se ve, para el desarrollo de la teoría y de la ciencia en general, ya que «un problema capital en la medición pedagógica es la escasa o inadecuada conceptualización del producto y de los procesos educativos, que se traduce en serias dificultades de medida... La necesidad del conocimiento preciso de los procesos y productos de la educación, sólo es posible con el apoyo en medidas correctas y adecuadas de las variables relevantes cuya identificación supone el desarrollo de teorías empíricamente validables» (Orden Hoz, 1985).

Es cierto, como señala Mateo (1988), que «las Ciencias de la Educación, como les ocurre a otras ciencias del ámbito socio-humano, no destacan por su riqueza en teorías fuertes, hecho que dificulta una aproximación axiomática al desarrollo de los procesos de medición».

Sin embargo, debe realizarse un esfuerzo de conceptualización teórica importante, para no incurrir en situaciones como aquellas a las que se ha llegado, en ocasiones, en la Psicología, donde desde posturas operacionalistas radicales «se ha llegado a hacer depender el concepto de inteligencia del proceso operacional de su medición a través de determinados tests. Llegándose a tantas definiciones de inteligencia como instrumentos independientes se hayan elaborado para su medición, invirtiéndose totalmente el proceso propuesto por la aproximación axiomática» (Mateo, 1988).

La validación de constructos y el desarrollo de la teoría están íntimamente vinculados, como se comprende, hasta el punto que Cronbach (1970) señala que el proceso de validación de un constructo exige partir de una teoría, a través de la cual se expliquen los resultados de un determinado instrumento. «La validación de un test es similar a la validación de cualquier teoría (...). La construcción misma de un test parte de una teoría.»

Expresado en otros términos, la elaboración y posterior validación de un constructo pasa por el establecimiento de lo que Cronbach y Meehl, en su clásico artículo de 1955, denominan la red nomológica, que no es otra cosa que un entramado sistema de leyes que constituyen la teoría en la que el constructo se inscribe.

Después de estas breves pinceladas introductorias sobre la importancia de la teoría en el desarrollo de la medición educativa, y del progreso científico de la Pedagogía misma, y la necesidad de elaborar marcos teóricos que expliquen los fenómenos que estudiamos, pasaremos a analizar con más detalle el concepto de validez haciendo una breve mención a sus diversos tipos para, en seguida, centrarnos en la validación de constructo.

2. CONCEPTO Y TIPOS DE VALIDEZ

El tema de la validez es complejo, controvertido y peculiarmente importante en la investigación. Mientras que es posible estudiar la fiabilidad de un instrumento de medida sin preguntarse por la naturaleza de lo que el instrumento mide, no es posible estudiar la validez sin preguntarse por el significado de lo medido (cfr. Kerlinger, 1975).

En el sentido más general podemos decir que un instrumento es válido si mide aquello que se pretende medir con él. Aunque de apariencia sencillo en la práctica es un concepto muy controvertido sobre el que se han suscitado numerosas polémicas entre los expertos.

Algunos autores señalan que se trata de una amalgama de consideraciones teóricas y prácticas que no acaban de separarse. Nunnally (1978), por ejemplo, señala que la validez tiene un significado excesivo, significados que es preciso diferenciar adecuadamente. Brinberg y Mac Grath (1982) en L meran 10 términos distintos de uso frecuente y distinguen 15 enfoques para el estudio de la validez. Messick (1980) enumera 17 términos, y en torno a los años setenta ya se podían encontrar hasta 40 términos sobre validez en la literatura.

Ya en fechas más recientes es conocida la controversia entre Ebel y Cronbach sobre el énfasis que el primero dice que hay que dar a la validez lógica o intrínseca, mientras que Cronbach

insiste en la necesidad de ahondar en la validez de constructo, con las oportunas pruebas experimentales.

Quizá el mejor modo de poner una cierta claridad en el estado de la cuestión sea acudir a una rápida revisión de las diversas ediciones de las Recomendaciones Técnicas de la American Psychological Association sobre los standards en la elaboración de tests (APA, 1954, 1966, 1974, 1985, 1986), que pretendieron poner un cierto orden y estabilizar las definiciones de validez.

La primera versión es de 1954, en ella quedan fijados los términos de validez de contenido, de criterio (concurrente y predictiva) y de constructo. Sin entrar en detalles que nos derivarían de nuestro objetivo señalaremos brevemente el significado de estos términos.

La validez de contenido, también denominada curricular, lógica o intrínseca, se refiere a la relevancia y representatividad del conjunto de ítems de un instrumento respecto a un universo temático previamente definido. Se trata de responder a la pregunta: ¿En qué grado esta prueba contiene una muestra representativa de elementos del universo que trata de explorar? Es, como se comprende, una característica intrínseca del instrumento en sí, y no de las inferencias que se pueden hacer con las puntuaciones derivadas del mismo.

La validez de criterio (criterion-oriented) se refiere a las relaciones de una determinada prueba con otras variables, medidas o criterios. Puede dividirse, como sabemos, en predictiva y concurrente, según la dimensión temporal que se establezca entre la prueba y el criterio.

La validez de constructo se refiere a un test cuando es interpretado como una medida de algún atributo o cualidad que no está «operacional-mente definido». El problema es responder a la pregunta: ¿Qué constructo es el responsable de la varianza de los resultados de este test? Un constructo es algún atributo postulado de las personas que se reflejará en los resultados de un test. «En la validación de tests el atributo respecto al que realizamos afirmaciones al interpretar el test es un constructo» (Cronbach y Meehl, 1955).

«El experto en medición, cuando investiga la validez de construcción de una prueba, generalmente desea conocer qué propiedad o propiedades psicológicas o de otra índole puede explicar la varianza de la prueba (...). Puede preguntar específicamente: ¿Mide esta prueba la capacidad verbal y la capacidad de razonamiento abstracto (...). Está investigando qué proporción de la varianza de la prueba es explicada por las construcciones: capacidad verbal y capacidad de razonamiento abstracto (...). Su interés es generalmente mayor por la propiedad que se mide que por la prueba misma» (Kerlinger, 1975).

El proceso de validar un constructo es, según Stenner, Smith y Burdick (1983), el proceso de adscribir significados a las puntuaciones obtenidas con un determinado instrumento, y es más necesario cuando aumenta el nivel de abstracción de la variable medida.

La segunda versión de las Recomendaciones es de 1966 —a la APA se unieron la AERA y el NCME, extendiéndose las consideraciones también al ámbito de lo escolar—, en ellas se aprecia la influencia que el trabajo de Crónbach y Meehl de 1955 tuvo, y que sigue siendo una referencia obligada. Mientras que en la primera versión la validez de constructo se entiende en su sentido más amplio, es decir, que no sólo se valida el constructo, sino la teoría misma en la

que el constructo se inscribe, de modo que es toda una red de relaciones la que se valida —no encajando bien, por ello, la validez de constructo dentro de los diversos tipos de validez—, en esta versión la atención se centra más en el constructo mismo.

Aunque ambos enfoques se complementan, el interés reside en validar el propio constructo, tal como lo mide un determinado instrumento. Y es así como entendemos hoy la validez de constructo. «El establecer varias hipótesis, una red nomológica, para verificar que estamos midiendo lo que queremos, es porque, en el caso de conceptos-abstracciones, no existe un único criterio claro de validación: no se valida la autoestima, por ejemplo, con una simple correlación con el rendimiento escolar, porque la autoestima, y cualquier constructo, implica más cosas» (Morales, 1988).

La tercer visión, de 1974, es importante porque deja muy claro que la validez no es una característica del instrumento de medida, sino de las interpretaciones y de las inferencias que se hacen con las puntuaciones del mismo. Decir que un instrumento es válido o no es un modo de decir coloquial, pero erróneo.

Por otra parte, la validez se reduce a dos tipos básicos: la validez relacionada con el significado o naturaleza del atributo medido (validez de constructo) y la validez relacionada con el uso del instrumento como indicador de otras variables o validez predictiva. Actualmente la validez de constructo se considera como el todo y la validez predictiva como una de sus partes, que ayuda a establecer relaciones entre el constructo y otras variables o constructos.

En la cuarta versión, de 1985, se realizan una serie de matizaciones sobre la validez de constructo y los procedimientos para su análisis, con mención expresa de la validación convergente y divergente, siguiendo la matriz multirrasgo-multimétodo de Campbell y Fiske (1959), sobre la que trataremos más adelante.

Como resumen de todo lo anterior, podemos señalar algunos puntos que nos parecen de interés:

1. Cuando decimos que un test es válido si mide lo que pretende medir podemos inducir a pensar que la validez es una propiedad del instrumento, lo cual no es correcto. Sería equivalente a decir que un test es válido o no lo es, cuando lo que ocurre es que un test puede ser válido para unos propósitos y no para otros.
2. La división de la validez en tipos, tal como sugieren en las normas de la APA, induce a confusión en el sentido de que pudiera pensarse que, con probar un tipo de validez el test ya está validado. No basta probar un tipo de validez para dar un test por válido (cfr. Messick 1980; Guion, 1980).
3. La validez lo es de las inferencias y no del test mismo, tal como, señala la tercera versión de la APA (1974): «La validez se refiere a la adecuación de las inferencias que se hacen de las puntuaciones de un test o de otras formas de evaluación».
4. La validez de constructo se entiende como el concepto más integrador de los diferentes tipos de validez. En efecto, se entiende un constructo como un concepto que representa una cualidad o atributo de los sujetos, de naturaleza más o menos abstracta, y que se supone

representado en las puntuaciones de un determinado instrumento, término que Cattell (1964) prefiere denominar concepto, ya que, a su juicio, une mejor la conceptualización teórica con la pura psicometría (aunque su propuesta no ha prevalecido).

Se trata, en cualquier caso, de desentrañar el rasgo o construcción subyacente a la varianza de las puntuaciones de un instrumento. El estudio de este tipo de validez requiere pruebas, evidencias experimentales, por lo que admite grados no reduciéndose a un índice o coeficiente. Es un proceso complejo, que requiere numerosos estudios, en cierto modo, siempre inacabado, como señalan Nunnally (1978) y Cronbach (1971).

Además, tampoco los datos cuantitativos son suficientes, requieren de una sólida argumentación lógica que los complementen. Como señala Messick (1980), los datos y la razón se armonizan e integran en una determinada interpretación.

Pasemos, ahora, a analizar los resultados de la aplicación de dos de los procedimientos más habituales en la validación de constructo.

3. LA VALIDACIÓN DEL CEED: PROCEDIMIENTOS Y RESULTADOS

Son dos los procedimientos que hemos empleado en la validación del Cuestionario para la evaluación de la Eficacia Docente (CEED): el análisis factorial clásico, pero con intención de comprobar determinadas hipótesis que veremos más adelante, y la matriz multirrasgo-multimétodo.

El análisis factorial se sitúa en el corazón de la validación de constructo. Como ya señalamos, la naturaleza compleja de los fenómenos que son objeto de nuestro interés en la investigación pedagógica requieren del recurso a técnicas de análisis que permitan estudiar las relaciones simultáneas entre un número elevado de variables, y se agrupan bajo la rúbrica: análisis multivariante.

Cuando el interés del investigador reside en llegar a obtener determinadas variables abstractas (latentes), no medidas, de naturaleza compleja, a través de una serie de variables simples empíricas, medidas, el análisis factorial es una de las herramientas de trabajo más poderosas con las que podemos contar, pero que ha de ser manejada con gran cautela, so pena de no obtener resultado significativo alguno.

Efectivamente, la aplicación del análisis factorial a conjuntos inespecíficos de variables medidas en un número mayor o menor de sujetos no puede producir fruto alguno. Esto significa que, aún utilizado en su enfoque más débil —no como técnica de análisis, sino como estrategia de investigación—el exploratorio, la contextualización teórica deberá ser lo más consistente posible, en lo que a las relaciones entre las variables se espera. Cuestión que en la validación de constructo es más que evidente.

En este punto queremos señalar, y en esto coincidimos con García Ramos (1986), que si el apoyo teórico es adecuado y la selección de las variables que intervendrán en el análisis también —el AF no descubre constructos, sino que revela o no su presencia—, aun fuera de estructuras causales su utilización puede considerarse confirmatoria, aunque no tenga la misma potencia probatoria que tiene en el seno de los modelos causales.

Cuestión esta última que tampoco deja de ser controvertida para algunos autores de la corriente «experimentalista», que, como sabemos, sólo aceptarían hablar de causalidad dentro del contexto propiamente experimental, por no mencionar a quienes no aceptan la idea de causa bajo ningún supuesto, cuestiones que ahora están al margen de nuestra exposición.

Veamos ahora la, utilización del análisis factorial en el estudio de la estructura del CEED, a partir de los datos obtenidos tras su aplicación a una muestra de alumnos de la Universidad de Navarra. Los detalles de la aplicación pueden consultarse en Marsh, Touron & Wheeler (1985).

Se trata de un instrumento de 62 ítems que combina nueve factores procedentes del SEEQ (véase Mars, 1982a, 1983, 1984), cuya estructura ha sido probada en numerosos estudios realizados en diversos contextos (cfr. Marsh, 1982b; Marsh & Hocevar, 1983), y los ítems de otro instrumento, el ENDEAVOR, con una estructura de siete factores, también probada en numerosas ocasiones (cfr. Frey, Leonard & Beatty, 1975). Estas dimensiones y su correspondencia son las que aparecen en el cuadro 1.

CUADRO 1 CORRESPONDENCIA ENTRE LOS FACTORES DEL "SEEQ" Y EL "ENDEAVOR"

Factores del SEEQ	Factores del ENDEAVOR
1. Aprendizaje/Valor	1. Logros de los alumnos
2. Interacción con el grupo	2. Discusión en clase
3. Raport individual	3. Atención personal
4. Exámenes/Calificaciones	4. Calificaciones
5. Carga de trabajo/Dificultad	5. Carga de trabajo
6. Organización/Claridad	6. Claridad de presentación
7. Entusiasmo	7. Organización/Planificación
8. Amplitud de enfoque	
9. Trabajos Extra-Clase/Lecturas	

Existen tres factores del SEEQ, como puede verse, que no tienen correspondencia con el ENDEAVOR: Entusiasmo del profesor, Amplitud de planteamiento y Trabajos extra-clase-lecturas.

Los ítems procedentes del SEEQ son 34 —si bien dos de ellos son globales y no están diseñados para definir ningún factor particular— y los del EN-DEAVOR 21. A ellos se han añadido siete ítems (A1 -A7) para definir mejor algunos factores según se indica en la tabla 1.

TABLA 1 - ANALISIS FACTORIAL CONJUNTO DE LOS ITEMS DEL "SEEQ" Y EL "ENDEAVOR"

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
APRENDIZAJE/VALOR										
14 M1. Clases intelectualmente estimulantes	34	27	25	-03	06	03	14	10	08	07
16 M2. Aprendido algo de auténtico valor	63	04	09	01	07	07	12	00	05	02
55 M3. Aumento del interés por la materia	47	22	06	06	05	09	15	03	08	-01
46 M4. Comprendido y aprendido la materia	70	01	02	15	00	04	-04	10	08	-08
15 F19. Comprendido los temas más complejos	53	09	15	17	04	05	-04	03	07	-13
32 F20. Aumento capacidad análisis temas relevantes	58	10	03	07	05	00	15	05	04	10
45 F21. Aumento de conocimientos/competencia	66	03	04	01	-01	14	12	10	07	10
ENTUSIASMO DEL PROFESOR										
2 M5. Profesor entusiasmado con la enseñanza	05	56	08	-01	07	13	20	-04	05	01
17 M6. Profesor dinámico en desarrollo curso	14	31	20	06	07	02	13	12	16	14
33 M7. Exposiciones atractivas por notas humor	07	49	03	04	18	09	08	05	03	07
47 M8. Estilo profesor mantiene interés clase	22	45	22	06	07	-01	00	10	01	04
62 A1. Profesor parece disfrutar con enseñanza	07	61	00	05	08	15	19	01	04	04
CLARIDAD DE EXPOSICION (ORGANIZACION)										
1-9. Explicaciones del profesor claras	16	24	50	19	02	-02	04	02	06	-01
19 M10. Temas bien preparados y explicados	13	22	50	21	-02	00	11	04	08	04
48 M12. Exposiciones facilitaron tomar notas	02	25	35	27	04	05	01	12	09	-09
9 F1. Explicaciones aclaraban material escrito	15	19	46	15	-01	04	13	07	05	03

(1) (2) (3) (4) (5) (6) (7) (8) (9) (10)

25 F2.	Explicación clara y síntesis principal	09	29	41	21	00	00	06	11	11	-02
39 F3.	Buen uso de los ejemplos	07	37	11	02	07	02	19	20	07	04
PLANIFICACION/OBJETIVOS (ORGANIZACION)											
60 M11.	Conocimiento metas y objetivos profesor	14	-07	22	58	02	-02	09	08	04	01
13 F13.	Clases bien preparadas	06	16	29	13	-11	12	30	09	-02	09
29 F14.	Se nos dio detallado programa del curso	-02	-11	08	58	16	02	09	10	03	14
43 F15.	Actividades clase programadas con orden	08	-04	25	44	-01	05	18	14	01	07
33 A2.	Desarrollo temas coincidente con programa	08	25	-09	50	12	04	08	10	13	-01
63 A3.	Profesor dio sus objetivos y criterios	14	15	-08	57	03	06	-05	10	16	01
INTERACCION CON EL GRUPO/DISCUSSION											
3 M13.	Discusión en clase estimulada	07	06	02	01	70	06	04	02	05	00
20 M14.	Se nos invitó a compartir ideas	03	06	05	02	66	09	15	07	08	-01
34 M15.	Se nos animó a preguntar	01	11	11	16	51	12	15	07	08	00
42 M16.	Se nos animó a expresar opiniones	00	08	-01	09	69	16	09	05	06	-02
12 F10.	Intervenciones en clase bien acogidas	04	04	21	03	44	23	15	08	03	00
28 F11.	Se nos animaba a intervenir en clase	06	08	-03	10	73	09	06	03	08	03
49 F12.	Compartir ideas/debatir las del profesor	03	07	-02	15	66	14	13	06	02	-01
RAPORT INDIVIDUAL/ATENCION PERSONAL											
4 M17.	Profesor amable en trato con alumnos	00	06	18	-06	29	39	09	12	08	-06
31 M18.	Acogidos al buscar ayuda/consejo	02	02	24	04	21	50	-01	17	09	00
36 M19.	Profesor interés por cada alumno	03	14	07	06	17	52	03	15	06	00
50 M20.	Profesor disponible a horas fijas	11	01	-04	07	-01	65	08	-04	11	05
11 F7.	Escuchaba problemas/dispuesto a ayudar	01	-03	23	-02	28	52	05	10	04	04
27 F8.	Tuve ayuda personal cuando la pedi	06	05	12	-01	09	60	09	11	06	04
41 F9.	Profesor interesado dificultades alumnos	-02	10	08	07	26	51	-01	13	10	04

(1) (2) (3) (4) (5) (6) (7) (8) (9) (10)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
AMPLITUD DE ENFOQUE										
5. M21. Contrastaba implicaciones teorías/enfoques	00	01	11	-02	12	01	67	03	04	01
21 M22. Origen ideas y conceptos desarrollados	05	15	07	08	10	-01	49	09	04	00
37. M23. Presentación de otros puntos de vista	-04	03	08	08	10	10	57	04	10	-03
51 M24. Discusión de los últimos avances materia	12	12	-14	09	-04	16	57	03	13	02
CALIFICACIONES/EXAMENES										
6 M25. Orientación posterior exámenes valiosa	03	-06	19	10	04	20	18	25	19	01
22 M26. Métodos evaluación justos y adecuados	01	04	09	05	06	04	04	77	01	-02
38 M27. Exámenes valoraban contenido adecuadamente	10	07	-09	18	-01	13	18	38	15	-07
61 F16. Calificaciones justas e imparciales	01	08	00	07	-02	11	03	78	03	01
30 F17. Calificaciones reflejaron rendimiento	05	-02	03	09	03	01	03	82	06	-02
44 F18. Calificaciones indicativas logros alumnos	08	01	04	09	04	01	05	77	02	00
TRABAJOS EXTRA-CLASE/LECTURAS										
7 M28. Lecturas/textos recomendados de calidad	07	-05	12	-11	07	-11	25	22	35	01
23 M29. Ayudaron comprensión asignatura	24	-01	04	06	13	-02	14	16	48	09
18 A4. Estimularon a profundizar asignatura	30	-02	08	-12	14	-02	13	07	43	01
53 A5. Bien integrados en el curso	-05	05	03	08	04	03	03	00	84	00
52 A6. Adecuados en dificultad y extensión	08	08	-08	18	-03	20	-01	00	58	10
54 A7. Relacionados con el trabajo de clase	03	06	03	12	-01	14	11	01	60	05
CARGA DE TRABAJO/DIFICULTAD										
57 M32. Dificultad asignatura (fácil-difícil)	-02	07	-05	02	04	-07	-01	-06	02	83
59 M33. Carga de trabajo (ligera-pesada)	-06	04	-06	05	-05	02	-01	-05	-02	86
58 M34. Ritmo de desarrollo (lento-rápido)	09	10	00	10	-06	-04	04	07	-04	49
10 F4. Trabajo duro en la asignatura	03	01	12	-02	03	01	00	04	10	83

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
26 F5. Preparación asignatura gran dedicación	00	-08	10	00	04	02	02	03	09	83
40 F6. Carga trabajo curso pesada	-02	-04	00	-02	02	05	02	02	03	83
ITEMS DE VALORACION GLOBAL										
8 M31. Valoración global del profesor	17	30	40	12	05	09	08	13	10	08
24 M30. Valoración global clases	12	29	40	17	02	08	07	10	12	05

Ambos instrumentos, según los numerosos estudios previos sobre ellos realizados, apoyan la multidimensionalidad del constructo que estamos estudiando (cfr. Frey, 1978; Marsh, 1981).

Los ítems procedentes de ambos instrumentos fueron analizados individualmente respecto a su pertinencia, su capacidad para diferenciar entre profesores «buenos», «medios» y «malos», y a su importancia. Análisis que están fuera de nuestro interés ahora (cfr. Marsh, Tournon & Wheeler, 1985).

La hipótesis planteada es que los 62 ítems del CEED miden 10 dimensiones de la eficacia docente —siete factores del ENDEAVOR complementados con sus correspondientes del SEEQ, más los tres factores del SEEQ que no tienen su correlato en el otro instrumento— (véase Marsh, 1981, quien llevó a cabo un trabajo similar en Australia). Para probar esta hipótesis se realizó un primer análisis factorial, con el que se pretendía mostrar la consistencia de las dimensiones. Es decir, que si se supone que determinados factores de ambos instrumentos miden lo mismo vendrán definidos, en la solución factorial, por el mismo factor. Primer análisis importante en la validación.

Como se puede apreciar en la tabla 1 las dimensiones hipotetizadas se reproducen con notable claridad. Con la excepción de los ítems F3 y F13, todos los demás ítems saturan en el factor previsto con mayor intensidad que en cualquier otro. La mayor parte de las saturaciones son mayores de 0,55, y sólo hay tres menores de 0,30. La mayor parte de las saturaciones de los ítems en los factores que no les corresponden teóricamente son menores de 0,1, el 95 por 100 son menores de 0,2 y menos del 1 por 100 son mayores de 0,30.

Se muestra, por ello, con bastante claridad la consistencia de las dimensiones hipotetizadas y, por tanto, la concordancia entre los factores de ambos instrumentos que miden las mismas dimensiones.

Además, y teniendo en cuenta el análisis de la matriz MRMM que presentaremos después, es necesario probar también, que las dimensiones de cada instrumento son consistentes por sí mismas. Para ello realizamos dos análisis factoriales separados, uno para el SEEQ y otro para el ENDEAVOR. Como se puede apreciar en la tabla 2, los factores del SEEQ se replican con notable claridad. Las saturaciones de cada ítem en los factores previstos es sustancial,

mientras que las saturaciones «fuera» del factor son mucho menores, escasamente alguna llega a 0,3.

TABLA 2 - ANALISIS FACTORIAL DEL "SEEQ"

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
APRENDIZAJE/VALOR									
M1. Clases intelectualmente estimulantes	33	42	05	06	08	09	08	13	03
M2. Aprendido algo de auténtico valor	69	08	00	09	08	05	-05	10	03
M3. Aumento del interés por la materia	49	25	01	06	13	14	01	02	-02
M4. Comprendido y aprendido la materia	68	-01	15	03	07	00	08	-04	-08
ENTUSIASMO DEL PROFESOR									
M5. Profesor entusiasmado con la enseñanza	06	45	05	13	14	18	-08	12	02
M6. Profesor dinámico en desarrollo curso	11	47	02	14	03	12	27	00	10
M7. Exposiciones atractivas por notas humor	03	42	04	18	19	14	-01	-04	06
M8. Estilo profesor mantiene interés clase	23	54	14	09	05	07	07	00	02
ORGANIZACION/CLARIDAD									
M9. Explicaciones del profesor claras	19	36	48	01	02	12	-05	07	00
M10. Temas bien preparados y explicados	17	35	44	01	01	14	06	07	06
M11. Conocimiento metas y objetivos profesor	27	-07	45	08	00	09	20	-09	09
M12. Exposiciones facilitaron tomar notas	07	30	40	08	08	10	12	-06	-09
INTERACCION CON EL GRUPO/DISCUSSION									
M13. Discusión en clase estimulada	05	02	-01	79	03	-03	-06	09	00
M14. Se nos invitó a compartir ideas	01	02	-02	80	00	13	07	-01	00
M15. Se nos animó a preguntar	04	05	14	59	11	14	06	00	02
M16. Se nos animó a expresar opiniones	02	-01	-01	85	08	08	03	-06	-02
RAPORT INDIVIDUAL									
M17. Profesor amable en trato con alumnos	-06	09	08	38	34	04	07	15	-05
M18. Acogidos al buscar ayuda/consejo	-04	10	15	28	48	-02	17	06	00

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
M19. Profesor interés por cada alumno	-01	09	09	19	61	03	08	07	04
M20. Profesor disponible a horas fijas	13	-01	-06	02	70	08	-02	-03	-01
AMPLITUD DE ENFOQUE									
M21. Contrastaba implicaciones teorías/enfoques	-05	05	04	08	-01	70	-01	16	02
M22. Origen ideas y conceptos desarrollados	08	09	09	12	00	48	05	13	01
M23. Presentación de otros puntos de vista	-02	-02	12	11	06	73	02	-04	-01
M24. Discusión de los últimos avances materia	19	04	-13	00	14	65	08	-07	01
CALIFICACIONES/EXAMENES									
M25. Orientación posterior exámenes valiosa	-01	01	21	04	27	16	31	14	03
M26. Métodos evaluación justos y adecuados	01	01	23	12	11	04	54	04	-07
M27. Exámenes valoraban contenido adecuadamente	12	01	06	03	20	19	45	-05	-08
TRABAJOS EXTRA-CLASE/LECTURAS									
M28. Lecturas/textos recomendados de calidad	12	06	-09	04	-06	12	38	44	14
M29. Ayudaron comprensión asignatura	32	-01	-02	12	10	12	21	34	06
CARGA DE TRABAJO/DIFICULTAD									
M32. Dificultad asignatura (fácil-difícil)	01	-04	02	04	-04	00	-07	05	89
M33. Carga de trabajo (ligera-pesada)	-07	-02	02	-07	08	00	-05	03	88
M34. Ritmo de desarrollo (lento-rápido)	08	18	-07	03	-09	02	23	24	47
ITEMS DE VALORACION GLOBAL									
M31. Valoración global del profesor	09	42	29	06	14	09	11	13	08
M30. Valoración global clases	18	40	30	04	12	09	09	10	06

Lo mismo cabría decir del análisis del ENDEAVOR que se recoge en la tabla 3, en la que se aprecian con claridad las siete dimensiones hipotetizadas para este instrumento.

Así pues, disponemos de evidencia suficiente sobre las dimensiones del constructo y de la semejanza entre las que se supone miden la misma dimensión. Además, el número de replicaciones que se han llevado a cabo en contextos muy diversos dan a estos resultados un grado de certidumbre más que razonable (cfr. la monografía de Marsh, 1987, en la que se recogen la mayor parte de los estudios realizados sobre el SEEQ e instrumentos similares). Además, análisis factoriales propiamente confirmatorios, realizados a partir de matrices multirasgo-multimétodo, han permitido comprobar la invariancia factorial de las dimensiones reflejadas (cfr. Marsh & Hocevar, 1984).

TABLA 3 - ANALISIS FACTORIAL DEL "ENDEAVOR"

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
CLARIDAD DE EXPOSICION							
F1. Explicaciones aclaraban material escrito	60	06	05	04	16jc	06	15
F2. Explicación clara y síntesis principal	63	03	02	08	15	11	10
F3. Buen uso de los ejemplos	50	08	13	11	-03	15	10
CARGA DE TRABAJO/DIFICULTAD							
F4. Trabajo duro en la asignatura	11	86	03	-02	01	01	03
F5. Preparación asignatura gran dedicación	-03	87	00	02	04	03	03
F6. Carga trabajo curso pesada	-03	85	-02	01	00	-03	-04
RAPORT INDIVIDUAL / ATENCION PERSONAL							
F7. Escuchaba problemas/dispueto a ayudar	06	01	64	20	06	07	02
F8. Tuve ayuda personal cuando la pedi	03	01	70	04	06	07	15
F9. Profesor interesado dificultades alumnos	05	02	48	30	08	12	03
DISCUSION EN CLASE							
F10. Intervenciones en clase bien acogidas	21	00	33	44	01	06	04
F11. Se nos animaba a intervenir en clase	01	04	07	81	06	01	06
F12. Compartir ideas/debatir las del profesor	05	-01	10	78	05	06	05
ORGANIZACION/PLANIFICACION							
F13. Clases bien preparadas	35	07	20	-08	30	14	11
F14. Se nos dio detallado programa del curso	-04	12	00	19	41	15	06
F15. Actividades clase programadas con orden	08	00	05	-01	81	04	03
CALIFICACIONES							
F16. Calificaciones justas e imparciales	05	01	14	-04	01	78	04
F17. Calificaciones reflejaron rendimiento	02	-01	-02	03	04	91	02
F18. Calificaciones indicativas logros alumnos	03	01	-02	05	08	83	04

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
LOGROS DE LOS ALUMNOS							
F19. Comprendido los temas más complejos	27	-12	-07	10	11	07	41
F20. Aumento capacidad análisis temas relevantes	02	05	09	06	06	01	79
F21. Aumento de conocimientos/competencia	02	06	05	00	00	08	83

Hasta aquí hemos visto cómo la aplicación del análisis factorial nos ha permitido encontrar evidencia matemática sobre los factores hipotetizados. Es en este sentido, fuera de los modelos usados, en el que señalábamos antes que el análisis factorial clásico puede considerarse confirmatorio.

Una vez que hayamos demostrado con suficiente evidencia las dimensiones del constructo, en este caso en ambos instrumentos, podemos calcular una nueva matriz P, a partir de la matriz de saturaciones factoriales, que ahora será de puntuaciones factoriales.

Así, multiplicando cada vector fila de la matriz en puntuaciones típicas (Z) por cada vector columna de la matriz de coeficientes factoriales (C), obtendremos un escalar: la puntuación del sujeto i en el factor n.

El paso siguiente es obtener una matriz de correlación entre los diferentes factores. En nuestro caso, la correlación entre las 16 dimensiones —nueve factores del SEEQ y siete del ENDEAVOR— nos permite obtener una matriz que, con ligeras variaciones, no sustantivas, es una matriz multirasgo-multimétodo.

En ella la dimensión multirasgo está constituida por las diversas dimensiones de la eficacia docente descritas anteriormente al presentar el análisis factorial; la dimensión multimétodo la constituyen los dos instrumentos utilizados: SEEQ y ENDEAVOR.

El enfoque multirasgo-multimétodo ha sido desarrollado por Campbell y Fiske (1959), siendo uno de los procedimientos que más difusión han tenido en el estudio de la validez de constructo, y que sigue siendo utilizado en la actualidad (cfr., p. ej., Marsh, 1987; Shavelson, Hubner & Stanton, 1976).

Este método integra simultáneamente el estudio de dos dimensiones de la validez de constructo denominadas convergente y divergente.

La validez convergente vendrá representada por las correlaciones entre las dimensiones que se supone miden el mismo constructo a través de métodos distintos (SEEQ y ENDEAVOR). Si métodos distintos verifican el mismo, o los mismos rasgos, tendremos más seguridad de que el constructo tiene realidad independiente del método con el que se mida (cfr. Thorndike, 1982).

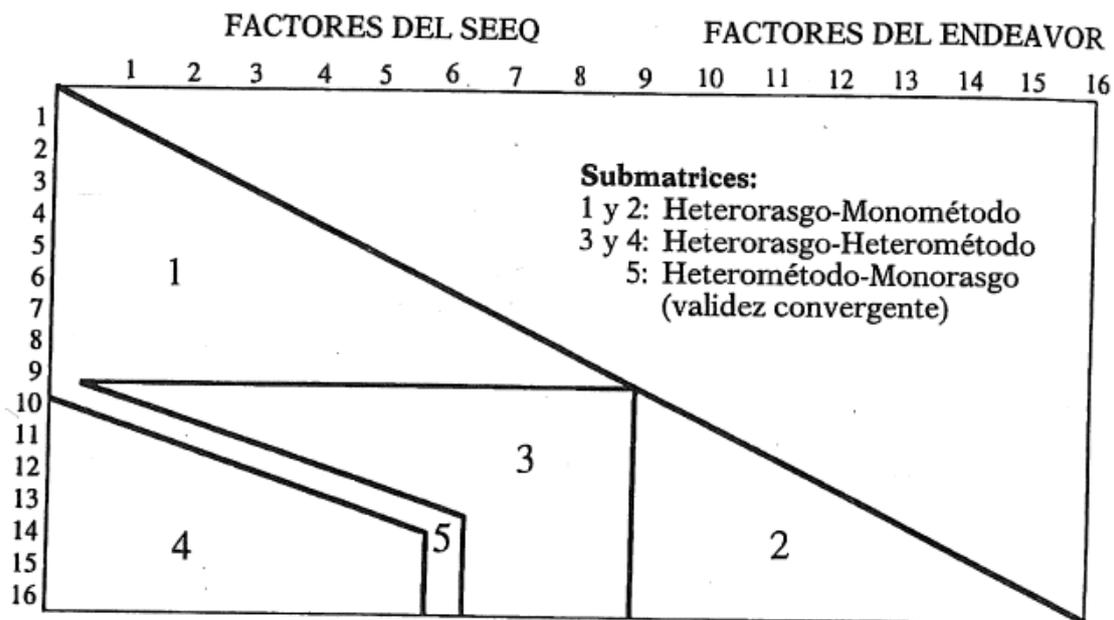
Por su parte, la validez divergente o discriminante vendrá apoyada por las relativamente bajas correlaciones entre rasgos diferentes medidos con el mismo método, lo que probará, en

mayor o menor grado, el carácter multidimensional del constructo, mostrando que cada dimensión está bien concebida en cuanto distinta.

Veamos, antes de pasar a los datos de los instrumentos que nos ocupan, cuál es la estructura de la matriz y cuáles son los criterios de convergencia y divergencia establecidos por Campbell y Fiske (1959). En el esquema de la matriz correspondiente a la validación del constructo que estamos estudiando, podemos distinguir: dos submatrices heterorasgo-monométodo (1 y 2), que contendrían los coeficientes de correlación entre rasgos distintos medidos, en cada caso, con el mismo método.

Por otra parte, las submatrices heterorasgo-heterométodo (3 y 4), que contendrían las correlaciones entre rasgos distintos medidos con métodos distintos. Finalmente, en la diagonal (5) se encontrarían los coeficientes de correlación entre los mismos rasgos medidos con métodos diferentes (monorasgo-heterométodo), que constituyen la evidencia acerca de la validez convergente.

ESQUEMA DE LA MATRIZ DE CORRELACION (MULTIRASGO-MULTIMETODO) ENTRE LOS FACTORES DEL SEEQ Y EL ENDEAVOR



Los criterios de análisis de la (1959), son los siguientes:

a) Criterio de convergencia

1. Los coeficientes de validez convergente, diagonal entre las submatrices 3 y 4 deben ser significativamente distintos de cero y suficientemente altos para poder considerar otros criterios ulteriores. FISKE (1982) analizando su método años después señala que este es el criterio fundamental.

b) Criterio de divergencia

2. Los coeficientes de validez convergente deben ser mayores que las correlaciones entre rasgos diferentes medidos con métodos distintos (submatrices 3 y 4). Como señala Morales (1988), lo menos que puede esperarse es que los coeficientes entre rasgos cuando no coinciden ni el método ni el rasgo sean menores que cuando el rasgo es el mismo aunque se mida con un método diferente. Desde el punto de vista práctico los coeficientes de validez convergente deben ser mayores que los de la fila y columna correspondiente (en las submatrices heterorasgo-heterométodo).
3. Los coeficientes de validez convergente deben ser mayores que los coeficientes de correlación entre rasgos distintos medidos con el mismo método (submatrices 1 y 2, heterorasgo-monométodo). En este caso se comparan los coeficientes de validez con los de las sub-matrices señaladas.
4. El último de los criterios se refiere a la estructura relacional (pattern) de las matrices que contienen los coeficientes de correlación entre rasgos diferentes medidos con el mismo método (submatrices 1 y 2). Sí la estructura relacional se mantiene es un criterio a favor de la semejanza entre constructos.

En la tabla 4 figuran los valores correspondientes a la matriz multirasgo-multimétodo entre los instrumentos SEEQ y ENDEAVOR. Veamos la aplicación de los criterios anteriores a dicha matriz.

1. Los coeficientes de validez convergente que figuran en la diagonal, en negrita, corresponden a los factores de ambos instrumentos que se supone miden lo mismo. Como se puede apreciar son sustancialmente altos y significativamente distintos de cero. Varían entre 0,71 y 0,93, por lo que satisfacen claramente el primer criterio.
2. El primer criterio de validez divergente requiere, como decíamos, que los coeficientes de validez convergente sean mayores que aquellos entre rasgos distintos medidos con métodos diferentes. La aplicación de este criterio requiere comparar cada uno de los siete valores de convergencia con los otros 14 coeficientes en su fila y columna. Este criterio se cumple en 97 de las 98 comparaciones, con lo que se satisface la condición con claridad.
3. Los coeficientes de convergencia deben ser mayores que los de las submatrices heterorasgo-monométodo. Su aplicación requiere que cada coeficiente sea comparado con los otros ocho coeficientes del SEEQ y los otros seis del ENDEAVOR. Este criterio se cumple en las 98 comparaciones.
4. La estructura relacional entre los factores del SEEQ deberá de ser similar a las correlaciones entre los factores del ENDEAVOR. Como ejemplo, señalemos que así como los factores 1 y 5 del SEEQ están altamente correlacionados, siguiendo el mismo patrón también deben estarlo el 10 y 14 del ENDEAVOR, como así ocurre. Una inspección de estas dos submatrices permite comprobar que el patrón de relaciones es similar en ambas, con lo que satisface también este criterio.

TABLA 4 - MATRIZ DE CORRELACION (MULTIRASGO-MULTIMETODO) ENTRE LOS FACTORES DEL SEEQ Y EL ENDEAVOR

Factores del SEEQ	Factores del ENDEAVOR															
	1	2	3	4	5	6	7	8	9							
1. Interacción con el grupo	(94)															
2. Aprendizaje/Valor	39	(92)														
3. Carga de trabajo/Dificultad	04	08	(79)													
4. Exámenes/Calificaciones	42	50	13	(85)												
5. Raport individual	68	43	-05	39	(90)											
6. Organización/Claridad	39	50	02	46	43	(91)										
7. Entusiasmo	47	65	22	40	43	64	(92)									
8. Amplitud de planteamiento	62	55	12	52	55	45	57	(89)								
9. Trabajos Extra-Clase/Lecturas	32	25	-05	26	29	18	24	36	(84)							
Factores del ENDEAVOR	Factores del SEEQ									Factores del ENDEAVOR						
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
10. Discusión en clase	93	37	03	38	69	38	43	59	29	(92)						
11. Logros de los alumnos	46	86	11	52	51	55	66	60	31	44	(87)					
12. Carga de trabajo	15	13	82	20	09	12	25	20	06	13	20	(91)				
13. Calificaciones	46	48	01	80	49	57	41	52	28	42	50	13	(94)			
14. Atención personal	75	37	08	44	81	40	48	56	32	72	46	18	45	(91)		
15. Claridad en la exposición	47	69	08	51	48	78	79	63	30	41	71	14	51	51	(89)	
16. Organización/Planificación	48	59	18	60	48	71	57	59	24	46	58	28	60	45	67	(78)

CONSIDERACIONES FINALES

Hasta aquí hemos visto dos de las principales vías de análisis de la validez de constructo, tanto desde una perspectiva teórica como aplicada. Conviene puntualizar, no obstante, que el proceso de validación ilustrado es parcial y que requiere de más evidencias empíricas, en concreto en nuestro país, pero la configuración del constructo mostrado parece estar fuera de cuestión. Ahora bien, como ya fue señalado, validar un constructo es un proceso inacabado, que requiere de un análisis múltiple: de interpretación y conceptualización del constructo en sí, del desarrollo de instrumentos de medida, de la recogida de datos y del razonamiento y reflexión teóricas.

Cada interpretación debe ser cautelosa y considerada una hipótesis tentativa que ha de ser sometida a prueba en diferentes contextos y con distintas aproximaciones metodológicas.

Este proceso supone la construcción de una red nomológica (Cronbach, 1971; Shavelson y cols., 1976) en la que se pongan en relación diferentes dimensiones del constructo bajo estudio y de éste con otros, en un proceso de elaboración teórica indispensable, a nuestro juicio, para el progreso del conocimiento pedagógico.

Finalmente, señalar que el proceso seguido se ha centrado en una intravalidación de un constructo (within-network) por el que tenemos un particular interés, ya que se inscribe en un ámbito de investigación que venimos cultivando desde hace algún tiempo.

Este es un paso previo para abordar otros estudios (between-network) en los que las evaluaciones de los alumnos se relacionen con otras variables externas, de criterio, que sirvan para determinar la importancia de las dimensiones estudiadas respecto a otros constructores esenciales dentro del desarrollo del proceso de enseñanza aprendizaje en la Universidad.

Deseamos terminar este trabajo con unas palabras de Thorndike del año 1913, correspondiente a la obra citada, que entendemos de plena vigencia setenta y seis años después: «Si aquellos que ponen objeciones al pensamiento cuantitativo en educación se pusieran a trabajar para comprenderlo; si aquellos que critican sus presupuestos y métodos hicieran realmente un trabajo experimental para mejorar su lógica y procedimientos; si aquellos que están ahora trabajando en la conceptualización y significado de la medida continuasen su trabajo, la próxima década traería consigo, con seguridad, un avance tanto de la teoría como de la práctica. De los avances de la década pasada podemos estar bien orgullosos».

BIBLIOGRAFIA

American Psychological Association (1954). «Technical Recommendations for Psychological Tests and Diagnostic Techniques». *Psychological Bulletin*, 51.

— (1966). *Standards for Educational and Psychological Tests and Manuals*. Washington, D. C., American Psychological Association.

- (1974). *Standards for Educational and Psychological Tests*. Washington, D. C., American Psychological Association.

- (1985). *Standards for Educational and Psychological Testing*. Washington, D. C., American Psychological Association.

- (1986). «Standards for Educational and Psychological Testing: Six Reviews». *Journal of Educational Measurement*, 23, 83-98.

Brinberg, G. W., y McGrath, J. E. (1982). «A Network of Validity Concepts within the Research Process». En Brinberg, D., y Kidder, L. H. (eds.). *Forms of Validity in Research*. San Francisco, Jossey-Bass.

Campbell, D. T., y Fiske, D. W. (1955). «Convergent and discriminant validation by the multitrait-multimethod matrix», *Psychological Bulletin*, 56, 81-105.

Cattell, A. B. (1966). *Handbook of Multivariate Experimental Psychology*. Chicago, Rand MacNally.

- (1964). «Validity and reliability: a Proposed more Basic Set of Concepts». *Journal of Educational Psychology*, 55, 1-22.

Comrey, A. L. (1985): *Manual de análisis factorial* Madrid, Cátedra.

Cooley, W. W., y Lohnes, P. R. (1971). *Multivariate Data Analysis*. New York, John Wiley & Sons.

Cronbach, L. J. (1970). *Essentials of Psychological Testing*. New York, Harper & Row Publ. Co.

- (1971). «Test validation». En Thorndike, R. L. (ed.). Educational Measurement. 2nd Ed. Washington, American Council on Education.

Cronbach, L. J., y Meelhl, P. (1955). «Construct validity in Psychological Tests». Psychological Bulletin, 52, 281-302.

Fiske, D. W. (1982). «Convergent-Discriminant Validation in Measurement and research Strategies». En Brinberg, D., y Kidder, L. H. (eds.). Forms of Validity Research, San Francisco, Jossey-Bass.

Frey, P. W. (1978). «A Two-dimensional Analysis of Student Ratings of Instruction». Research in Higher Education, 9, 69-91.

Frey, P. W.; Leonard, D. W., y Beatty, W. W. (1975). «Student Ratings of Instruction: Validation Research». American Educational Research Journal 12, 327-336.

García Ramos, J. M. (1986). «Validación de constructo en el ámbito pedagógico». Revista Española de Pedagogía, 174, 535-554.

Guión, R. M. (1980). «On Trinitarian Doctrines of Validity». Professional Psychology, 11, 385-398.

Guttman, L. (1971). «Measurement as Structural Theory». Psychometrika, 36, 329-347.

Harman, H. H. (1967). Modern Factor Analysis. Chicago, The University of Chicago Press.

Hope, K.(1972). Métodos de análisis multivariado. Madrid, Instituto de Estudios Políticos.

Jones, L. V. (1971). «The Nature of Measurement». En Thorndike, R. L. (ed.). Educational Measurement. Washington, American Council on Education.

Jóreskog, K G. (1974). «Analizing Psychological Data by Structural Analysis of Cova-riance Matrices». En Krantz, D. H.; Atkinson, R. C.; Luce, D., y Suppes, P. (eds.) Contemporary Developements in Mathematical Psychology (vol. 2). San Francisco, Freeman.

- (1977). «Structural Equation Models in the Social Sciences: Specification, Estimation and Testing». En Krishnaiah, P. R. (ed.). Applications of Statistics. Amsterdam, North-Holland.

Jóreskog, K. G., y Sórbom, D.(1976). COFAMM. Confirmatory Factor Analysis with Model Modification. User's Guide. Chicago, National Educational Resources.

- (1983). LISREL VI. Analysis áf Linear Structural Relationships by Maximum Likelihood and Least Square Methods. Chicago, International Educational Services.

Kerlinger, F. (1975). Investigación del comportamiento. México, Interamericana.

Krantz, D. E, y Tversky, A. (1971). «Conjoint-measurement Analysis of Composition Rules in Psychology». Psychological Review, 78, 151-169.

López Feal, R. (1986). Construcción de Instrumentos de Medida en Ciencias Conductuales y Sociales. Vol. I, Barcelona, Alamex.

Marsh, H. W. (1981). «Students' Evaluations of Tertiary Instruction: The Applicability of American Surveys in an Australian Setting». *Australian Journal of Education*, 25, 177-192.

- (1982a). «SEQ: A Reliable, Valid, and Useful Instrument for Collecting Students' Evaluations of University Teaching». *British Journal of Educational Psychology*, 52, 77-95.
- (1982b). «Validity of Students' Evaluations of College teaching: A Multitrait Multimethod Analysis». *Journal of Educational Psychology*, 74, 264-279.
- (1983). «Multidimensional ratings of teaching Effectiveness by Students from Different Academic Settings and their Relationship to Student/course/Instructor Characteristics. *Journal of Educational Psychology*, 75, 150-166.
- (1984). «Students' Evaluations of University teaching: Dimensionality, Reliability, Validity, Potential biases, and Utility». *Journal of Educational Psychology*, 76, 707-754.
- (1987). «Students' Evaluations of University teaching: Research Findings, Methodological Issues, and Directions for Future Research». *International Journal of Educational Research*, 11(3), 255-379.

Marsh, H. W., y Hocevar, D. W. (1983). «Confirmatory Factor Analysis of Multitrait-Multimethod Matrices». *Journal of Educational Measurement*, 20, 231-248.

- (1984). «The Factorial Invariance of Student Evaluations of College Teaching». *American Educational Psychological Journal* 21(2), 341-366.

Marsh, H. W.; Touron, J., y Wheeler, B. (1985). «Students' Evaluations of University Instructors: The Applicability of American Instruments in a Spanish Setting». *Teaching & Teacher Education*, 1(2), 123-138.

Mateo, J. (1988). «Medición educativa. Estado de la cuestión en el ámbito español». En Dendaluce, L (coord.). *Aspectos metodológicos de la investigación educativa*. Madrid, Narcea.

Messik, S. (1980). «Test Validity and the Ethics of Assessment». *American Psychologist*, 35, 1012-1027,

Morales, P. (1988). *La medición de actitudes en Psicología y Educación*. San Sebastián, Ed. Tarttalo, S.A.

Mulaik, S. A. (1972). *The Foundations of Factor Analysis*. New York, McGraw-Hill. Nunnally, J. C. (1978). *Psychometric Theory*. New York, McGraw-Hill.

Orden Hoz, A. (1985). «Medida». *Diccionario de Ciencias de la Educación*. Madrid, Anaya.

Shavelson, R. J.; Hubner, J. J., y Stanton, G. C. (1976). «Self-concept: Validation of Construct Interpretations». *Review of Educational Research*, 46, 407-441.

Stenner, A. J.; Smith, M., y Burdick, D. S. (1983). «Toward a Theory of Construct Definition». *Journal of Educational Measurement*, 20, 305-316.

Stevens, S. S. (1951). Mathematics, Measurement and Psychophysics. Handbook of Experimental Psychology. New York, Wiley.

Thorndike, E. L. (1913). An Introduction to the Theory of Mental and Social Measurements. Tomado del capítulo «Measurement» by E. L. Thorndike. En Broudy, H. S. y cols. (1973). Philosophy of Educational Research, New York, John Wiley & Sons, Inc.

- (1982). Applied Psychometrics. Boston, Houghton-Mifflin.

Yela, M.(1965). «Necesidad y posibilidad de la medida en Psicología». En Psicometría y Estadística. Madrid.

RESUMEN

En el presente trabajo se analiza la importancia del desarrollo de la teoría en la construcción del saber pedagógico, proceso que -dentro de los procesos de medida implicados en la investigación experimental- requiere prestar una mayor atención al concepto de validez, principalmente a la validez de constructo, que constituye el aspecto más integrador de los distintos tipos de validez.

Por otra parte se analiza la aplicación del análisis factorial clásico y de la matriz multirasgo-multimétodo al estudio de la validez de constructo de un instrumento para la evaluación de la eficacia docente. La aplicación de ambos procedimientos muestra con claridad la multidimensionalidad del constructo estudiado.

SUMMARY

The importance of theoretical background that supports measurement instruments used in educational research is discussed in this paper.

The integration of the theory in the measurement process is the most accurate way to approach the complexity of phenomena of interest in education. We argue that much more attention must be paid to the validity of measurement instruments, mainly to construct validity, the most integrative of validity concepts.

On the other hand, classical factor analysis and the Multitrait-Multimethod matrix, as proposed by Campbell & Fiske (1959), are used to study the construct validity of an instrument designed to evaluate teaching effectiveness. Both methods support the multidimensionality of the construct under study and replicate similar findings from Australian Studies of the same instrument.