



Revista Escuela de Administración de
Negocios

ISSN: 0120-8160

investigaciones@ean.edu.co

Universidad EAN
Colombia

Pascual Gómez, Isabel

Modelos jerárquicos lineales: Un caso de aplicación en la Universidad

Revista Escuela de Administración de Negocios, núm. 64, septiembre-diciembre, 2008, pp. 55-66

Universidad EAN

Bogotá, Colombia

Disponible en: <http://www.redalyc.org/articulo.oa?id=20612981005>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal

Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

MODELOS JERÁRQUICOS LINEALES:

UN CASO DE APLICACIÓN EN LA UNIVERSIDAD

Isabel Pascual Gómez*

RESUMEN

Este artículo aborda la comprobación del los índices de validez de un cuestionario de evaluación docente en la Universidad, a través de la utilización de Modelos Jerárquicos Lineales.

Los resultados obtenidos sugieren que este tipo de evaluaciones deben interpretarse con cautela especialmente si van a ser utilizadas con fines sumativos por los responsables académicos de la Universidad.

PALABRAS CLAVE

Modelos jerárquicos lineales
Eficacia docente
Índices de validez

1. INTRODUCCIÓN

La evaluación de la docencia de los profesores a través de encuestas de opinión a los alumnos se apoya en un modelo multidimensional de lo que debe ser **un buen profesor, un profesor competente** sobre el cual el alumno puede proporcionar información relevante.

Los estudios consultados sostienen que las encuestas, con sus virtudes y defectos, son el procedimiento que tiene tras de sí más estudio, más análisis científico, y más soporte metodológico de resultados contrastados. Esto no implica que sea un procedimiento sin contradicciones ni limitaciones técnicas.

Metodológicamente los estudios han trabajado intentando demostrar la dimensionalidad

* Profesora de Métodos de Investigación en Educación de la Universidad de Alcalá de Henares. Doctora en Ciencias de la Educación por la UCM, Licenciada en Filosofía y Ciencias de la Educación-Sección Ciencias de la Educación, por la Universidad Complutense. Su área de investigación se centra en la evaluación de la calidad en la Universidad, especialmente en la investigación de la evaluación de la eficacia docente.

del concepto eficacia docente, comprobando la fiabilidad entendida como consistencia interna en los cuestionarios y asegurando la validez de estas encuestas a través de: coeficientes de correlación, análisis de regresión y análisis factorial entre otros procedimientos estadísticos. (Cohen 1980; Marsh 1987; Feldman 1996; Fernández 1992; Tejedor 1988; Villa 1993; Apodaka 1999).

Para comprobar su validez, suelen seguirse dos caminos complementarios, la validación convergente y la validación divergente.

El procedimiento que aborda el problema de la validez de la evaluación del profesor intentando comprobar que existe una relación apreciable entre la evaluación de los alumnos y una serie de criterios generales que están estrechamente relacionados con eficacia docente (rendimiento de los alumnos, satisfacción del alumno con el profesor, autoevaluaciones....) se conoce como validación convergente.

La validación divergente comprueba que estas evaluaciones no están relacionadas con variables o indicadores no relacionados con calidad o eficacia docente como: número de alumnos por clase, dificultad de la asignatura, etc.

Este método de validación a pesar de estar muy extendido, ha sido criticado por numerosas investigaciones en sus limitaciones técnicas.

ABSTRACT

This article describes the checking up of validity rates of a teacher evaluation poll applied at EAN University through the use of hierarchical lineal models.

The results of this study suggest that the results of this kind of evaluation should be carefully interpreted if being used with summative purposes by the university academic authorities.

KEY WORDS

*Hierarchical Lineal Models
Teaching effectiveness
Validity rates.*

2. OBJETIVO

El objetivo de este trabajo es ofrecer una alternativa al análisis de validez convergente tradicional de este tipo de instrumentos mediante una adaptación de este procedimiento a los modelos jerárquicos lineales.

3. METODOLOGÍA GENERAL

Para llevar a cabo la determinación de la validez del instrumento se aplicó esta metodología a un cuestionario de evaluación docente de la Universidad Antonio de Nebrija. El instrumento elaborado por la Universidad es una escala de elección múltiple de 20 ítems agrupados en torno a cuatro factores, obtenidos a través de un análisis factorial: Competencia del profesor, Capacidad de interacción del profesor, Medios y Materiales y Cumplimiento formal. Trabajaremos en la validación de los dos primeros.

Los datos de la muestra pertenecen a las evaluaciones semestrales realizadas por los alumnos de forma anónima. En las Tablas 1 y 2 se describen las poblaciones de alumnos y profesores evaluados.

TABLA 1
Tabla de alumnos

Periodo	Total
Primer Semestre Curso 97-98	3.045
Segundo Semestre Curso 97-98	2.348
Primer Semestre Curso 98-99	3.874
Segundo Semestre Curso 98-99	2.348
Total	11.615

TABLA 2
Tabla de profesores evaluados

Periodo	Total
Primer Semestre Curso 97-98	159
Segundo Semestre Curso 97-98	170
Primer Semestre Curso 98-99	235
Segundo Semestre Curso 98-99	174
Total	738

Medir la percepción del alumno sobre la eficacia docente es relativamente sencillo. Si lo que se pretende es que esa información sea válida, es decir que mida lo que dice medir, el procedimiento se complica.

Este trabajo pretende validar si se mantiene la relación existente entre la satisfacción del alumno con la docencia recibida con los predictores Competencia e Interacción docente si se incorpora al estudio el agrupamiento jerárquico o anidado de los datos.

3.1 Descripción de la Lógica de los modelos Jerárquicos Lineales

Los modelos jerárquicos consideran que las agrupaciones en los datos son más similares que las observaciones entre grupos distintos y que es necesario contrastar la posible dependencia o ausencia de independencia entre dichas observaciones. Para más información sobre modelos jerárquicos lineales puede consultarse a Bryk y Raudenbush (1992), Goldstein (1995), y Bosker y Snijders (1999).

La formulación de este tipo de modelos se realiza a través de sistemas de ecuaciones de regresión jerárquicas y se centran en una estimación más completa de los coeficientes

de regresión (β) y en una estimación más numerosa de los componentes de varianza (μ) como describiremos a continuación.

Un modelo básico de dos niveles y su correspondiente sistema de ecuaciones de regresión jerárquicas sería como el que aparece a continuación:

Nivel 1

$$Y_{ij} = B_{0j} + B_{1j}x_{ij} + e_{ij}$$

Nivel 2

$$B_{0j} = B_{00} + B_{01}z_j + \mu_{0j}$$

$$B_{1j} = B_{10} + B_{11}z_j + \mu_{1j}$$

El modelo se describe por los siguientes componentes:

- ♦ Dos niveles de agrupamiento. En este modelo el nivel micro o nivel 1 representa el nivel inferior de la jerarquía (alumno) considerándose el nivel macro o nivel superior de la jerarquía el nivel 2 (en nuestro caso el nivel profesor).
- ♦ Una variable dependiente regresora o variable a estudiar (Y_{ij}) en nuestro caso la satisfacción del el alumno con el profesor.
- ♦ Diversas variables independientes o predictoras que vendrían representadas por los factores objeto de validación.
- ♦ Varios coeficientes de regresión. Los coeficientes B_0 y B_1 son los interceptos y pendientes de un procedimiento normal de regresión pero incorporando en cada uno de ellos un subíndice, Permitiéndonos estos subíndices contemplar la posibilidad

de que los interceptos y las pendientes varíen entre las unidades de nivel superior.

- ♦ Diversos componentes de varianza. Estos componentes explicarán la variación atribuible a los dos niveles de la ecuación.
- ♦ El estudio de la covarianza o posible existencia de relación de las variables entre las diversas jerarquías.

4.

MODELO DE VALIDACIÓN CONVERGENTE DE LA EFICACIA DOCENTE

El modelo jerárquico propuesto por este trabajo pretende conocer si se puede confirmar la relación existente entre la satisfacción del alumno, y dos predictores de eficacia tradicionalmente aceptados: Competencia docente e Interacción docente.

El modelo observa los datos desde distintos niveles de agrupamiento y estudia el comportamiento con variables que pasaremos a detallar a continuación.

Los niveles de anidamiento o agrupación de datos que forman parte del modelo son:

- ♦ **Nivel 1 o nivel alumno.** Se va utilizar una variable de identificación del alumno que asigna un código a cada uno de los alumnos.
- ♦ **Nivel 2 o nivel profesor.** Supone la creación de una variable de identificación por profesor.

- ♦ La variable dependiente del modelo o satisfacción del alumno con el profesor.
- ♦ Las variables independientes están centradas en torno a la media de las unidades de nivel 2, es decir, centradas en torno a la media de satisfacción del alumno con el profesor. Esto supuso la recodificación de las variables, las que pasaron por varias etapas, hubo que calcular las puntuaciones en los dos factores competencia e interacción para cada alumno, la media en los dos factores de cada profesor y una vez obtenidas estas hallar las diferencias individuales entre la puntuación en el factor y la media de cada profesor.

El modelo se acoge a los supuestos mencionados en el modelo anterior:

- ♦ Los errores son aleatorios, normales e independientes.
- ♦ La varianza de los residuos es homogénea.
- ♦ La media de los interceptos es una constante.

El objetivo es conocer para cada periodo académico:

- ♦ La media general de satisfacción.
- ♦ Los incrementos diferenciales que se producen en interacción y en competencia docente y como afectan a la satisfacción.
- ♦ La varianza explicada por alumnos y profesores.
- ♦ La relación existente entre estos dos constructos entre sí y con la variable dependiente.

Las relaciones existentes se formulan como aparecen a continuación:

Ecuación en nivel alumno o nivel 1.

$$y_{ij} = \beta_{0ij} + \beta_{1j} x_{1ij} + \beta_{2j} x_{2ij} \quad \text{Ecuación 4.1}$$

Donde:

y_{ij} - Satisfacción del alumno i con el profesor j.

β_{0ij} - Media de satisfacción con ese profesor.

β_{1j} - Incremento que se produce en la satisfacción del alumno al aumentar un punto la competencia docente del profesor.

x_{1ij} - Diferencia entre la competencia docente media de un profesor y la puntuación de alumno en ese constructo.

β_{2j} - Incremento que se produce en la satisfacción del alumno al aumentar un punto la capacidad de interacción docente del profesor.

x_{2ij} - Diferencia entre la interacción docente media de un profesor y la puntuación del alumno en interacción.

Ecuación en el nivel 2 o nivel profesor

$$\begin{aligned} \beta_{0ij} &= \beta_0 + \mu_{0ij} + \epsilon_{0ij} \\ \beta_{1j} &= \beta_1 + \mu_{1j} \\ \beta_{2j} &= \beta_2 + \mu_{2j} \end{aligned} \quad \text{Ecuación 4.2}$$

Donde:

β_{0ij} - es la satisfacción del alumno i con el profesor j.

β_0 - Es la media de Satisfacción general.

μ_{0j} - Error cometido por el alumno al valorar su satisfacción con el profesor j.

ε_{ij} -	Error cometido por el alumno en el proceso de medida.	B_2 -	Media de interacción docente.
B_{1j} -	Competencia docente de un profesor	μ_{2j} -	Error cometido al valorar la capacidad de interacción docente del profesor j.
B_1 -	Media de competencia docente.		
μ_{1j} -	Error cometido al valorar la competencia docente del profesor j.		
B_{2j} -	Interacción docente de un profesor.		

Los resultados de los interceptos obtenidos y la matriz de varianzas y covarianzas se muestran en la Tabla 3.

TABLA 3
Resumen de los interceptos obtenidos con el modelo de validación convergente

	Satisfacción General	Incrementos en competencia docente	Incrementos en Interacción docente
97-1c	5,263 (0,065)	0,805(0,027)	0,269(0,024)
97-2c	5,276 (0,082)	0,754 (0,033)	0,291(0,027)
98-1c	5,188(0,072)	0,831(0,028)	0,188 (0,025)
98-2c	5,338(0,084)	0,332(0,056)	0,410(0,043)

Los parámetros fijos resultan todos significativos ya que si dividimos el valor del estimado por su error típico que aparece a continuación el cociente resultante es siempre superior a 2, que es el valor utilizado para el nivel de significación del 0.05. Por lo tanto, las 2 variables utilizadas competencia e interacción docente son predictoras de la satisfacción.

Los intervalos de confianza de la variable satisfacción general serían:

5,263 \pm 1,96*0,065 (5,383 5,143)
 5,276 \pm 1,96*0,082 (5,436 5,116)
 5,188 \pm 1,96*0,072 (5,32 5,048)
 5,330 \pm 1,96*0,084 (5,49 5,166)

Las puntuaciones se han mantenido estables en el tiempo obteniéndose una satisfacción general alta, (valores ligeramente superiores a 5 sobre un máximo de satisfacción de 7).

Los intervalos de confianza para los constructos competencia e interacción docente serían:

Competencia docente

0,805 \pm 1,96*0,027 (0,857 0,752)
 0,754 \pm 1,96*0,033 (0,818 0,689)
 0,831 \pm 1,96*0,028 (0,885 0,776)
 0,332 \pm 1,96*0,056 (0,224, 0,441)

Interacción con el alumno

$0,269 \pm 1,96*0,024$ (0,316 0,221)
 $0,291 \pm 1,96*0,0291$ (0,343 0,238)
 $0,188 \pm 1,96*0,025$ (0,237 0,139)
 $0,410 \pm 1,96*0,045$ (0,321,0,498)

Se produce una mayor valoración de la competencia docente del profesor que de su capacidad de interacción con el alumno. Por cada punto que aumentemos la competencia docente, la media de satisfacción se incrementaría entre 0,33 y 0,8 puntos.

Por cada punto que aumentemos la capacidad de interacción con el alumno la media de satisfacción se incrementaría entre 0,18 y 0,41 puntos.

En vista de estos resultados podría afirmarse que tiene mayor peso en la satisfacción del alumno con el profesor, su competencia docente que su capacidad de interacción.

Los componentes aleatorios para el modelo se resumen en la Tabla 4.

Respecto a la variable satisfacción:

En el nivel 1 o nivel alumno.

TABLA 4
Varianzas obtenidas en el nivel 1 en el modelo de validación convergente

	Var e_{oij}	Error
97-1c	0,839	0,023
97-2c	0,834	0,026
98-1c	0,689	0,017
98-2c	0,981	0,029

Las varianzas asociadas a los alumnos son significativas. El modelo explicaría la diferencia entre las medias de los alumnos en satisfacción docente.

En el nivel 2 o nivel profesor la Tabla 5 muestra las varianzas asociadas a las medias de satisfacción

TABLA 5
Varianzas obtenidas en el nivel 2 en el modelo de validación convergente

	Var μ_{oj2}	Error
97-1c	0,576	0,074
97-2c	1,034	0,123
98-1c	1,139	0,111
98-2c	1,122	0,130

Las varianzas del intercepto también son significativas. También se explicarían a este nivel las diferencias entre las medias obtenidas por los profesores en satisfacción docente.

Respecto a las pendientes o los incrementos que se han producido en la satisfacción al aumentar la competencia e interacción docente, los resultados aparecen en la Tabla 6.

TABLA 6
Varianzas obtenidas en factor competencia docente en el modelo de validación convergente

	Var μ_{1j2}	Error
97-1c	0,017	0,010
97-2c	0,025	0,015
98-1c	0,053	0,014
98-2c	0,261	0,053

Los resultados obtenidos muestran unas pendientes de signo positivo. Esto indicaría en caso de probarse su significatividad, que con los profesores cuyos alumnos tienen un nivel medio de satisfacción, el nivel de satisfacción de los alumnos aumenta cuando se incrementa su percepción de que el profesor es competente.

Con los resultados obtenidos no podemos probar la significatividad de las pendientes en competencia docente. La Tabla 7 resume las varianzas de la pendiente interacción con el alumno.

TABLA 7
Varianzas obtenidas en el factor interacción docente en el modelo de validación convergente

	Var μ_{2j}	Error
97-1c	0,013	0,008
97-2c	0,010	0,009
98-1c	0,049	0,012
98-2c	0,131	0,031

Los resultados obtenidos muestran pendientes de signo positivo. Esto indicaría en caso de probarse su significatividad que con los profesores cuyos alumnos tienen un nivel medio de satisfacción, el nivel de satisfacción de los alumnos aumenta cuando aumenta su

percepción de que el profesor tiene una buena relación y comunicación con los alumnos.

Los resultados obtenidos no pueden probar la significatividad de las pendientes en interacción en los alumnos en todos los ficheros de datos.

Nuestro modelo no podría explicar los incrementos que se producen en la satisfacción del alumno si consideramos como predictores la competencia de un profesor y su capacidad de interacción con el alumno.

Análisis de covarianzas

Volvemos a realizar el cálculo de la covarianza con el objetivo de contrastar los resultados anteriores y comprobar la relación entre la satisfacción del alumno y los predictores. Se obtienen valores bajos y de signo negativo.

Una covarianza de signo negativo mostraría que con los profesores cuyos alumnos tienen un nivel medio bajo de satisfacción, el nivel de satisfacción de los alumnos se incrementaría poco cuando aumenta su percepción de que el profesor es competente.

Los resultados del contraste estadístico que aparecen en la Tabla 8 no han podido demostrar su significatividad.

TABLA 8
Resumen de la matriz de varianzas y covarianzas en el modelo de validación convergente

	Var μ_{10j}	Error μ_{10}	Var μ_{20j}	Error μ_{20}	Var μ_{21j}	Error μ_{21}
97-1c	-0,016	0,021	-0,018	0,019	-0,014	0,08
97-2c	-0,032	0,034	-0,005	0,028	-0,015	0,011
98-1c	-0,050	0,030	-0,035	0,027	-0,045	0,012
98-2c	-0,249	0,064 (-4)	0,030	0,047	-0,130	0,038

Como en el caso anterior los valores de la covarianza entre satisfacción con el profesor y la capacidad de interacción con el alumno son muy bajos y de signo negativo. El nivel de satisfacción con el profesor aumenta poco al incrementar la percepción del alumno sobre la capacidad de interacción del profesor.

Respecto a la relación entre los dos predictores, competencia docente e interacción con el alumno no ha podido tampoco probarse una relación de dependencia entre ellos.

Las varianzas asociadas a los dos niveles no son todas significativas. Este modelo explicaría gran parte de la varianza entre las medias de los alumnos y de los profesores en satisfacción docente, pero no podría explicar los incrementos que se producen en la satisfacción del alumno al variar la competencia e interacción docente atribuidas a sus profesores.

Según Bryk y Raudenbush (1992) "Un valor positivo de T01 y T02 indicarían que un profesor

con un alto valor en satisfacción tiende a tener una pendiente alta". Es decir, en nuestro caso un alto incremento en competencia e interacción docente. No se cumple ninguna de esas condiciones, son valores bajos, de signo negativo y no significativos.

No se podría afirmar con este modelo que en las medias obtenidas en satisfacción docente tenga mayor peso para un alumno la competencia docente que la capacidad de interacción con su profesor.

Para contrastar estos resultados se van a calcular los coeficientes de correlación entre el intercepto y la pendiente. Para el cálculo de este coeficiente utilizaremos la siguiente ecuación:

$$\rho(\beta_{0j}, \beta_{1j}) = \tau_{01} / (\epsilon_{00} \tau_{11})^{1/2} \quad \text{Ecuación 4.3}$$

Los resultados obtenidos son los que aparecen en la Tabla 9:

TABLA 9
Operaciones intermedias para el cálculo del
coeficiente de correlación entre
la competencia docente y la satisfacción

T00	T01	T11
0,576	-0,016	0,017
1,034	-0,032	0,025
1,139	-0,05	0,053
1,121	-0,272	0,28

$TOO * T11$	$TOO * T11^{1/2}$	$TO1 / (TO0 T11)^{1/2}$
0,009792	0,09895454	-0,161690417
0,02585	0,16077935	-0,199030532
0,060367	0,24569697	-0,203502715
0,31388	0,56024994	-0,485497594

Los coeficientes de correlación obtenidos entre la satisfacción y la competencia docente están en la mayoría de los casos por debajo de 0,30 lo que indica una correlación débil, y son de signo negativo, lo cual corrobora los resultados obtenidos en otras investigaciones que hallaron que los profesores más competentes con sus alumnos no son por ello menos valorados sino al contrario.

Los coeficientes de correlación obtenidos entre la satisfacción docente y la capacidad de interacción tampoco muestran una relación fuerte y positiva entre la variable y el predictor.

Los resultados están en la misma línea del apartado anterior, en el cual tras el cálculo de la covarianza no podríamos afirmar que estos

dos constructos competencia e interacción, sean totalmente separables y evaluables por el alumno a la hora de expresar su opinión sobre la eficacia y satisfacción con un profesor determinado.

6. CONCLUSIONES

Respecto a la validez del instrumento

Los resultados obtenidos nos hacen suponer que la elección de los predictores como criterios de validez de la satisfacción docente (Competencia del profesor y Capacidad de interacción o comunicación con el alumno) ha sido acertada.

Tanto las medias como las varianzas en competencia e interacción docente han resultado estadísticamente significativas.

Respecto al grado de influencia de cada uno de los predictores en la satisfacción del alumno los resultados no son tan satisfactorios. No podemos afirmar que uno de ellos tenga más peso que el otro en la satisfacción de un alumno, ni que exista una relación de dependencia entre ellos.

Los resultados nos permiten afirmar que el alumno a la hora de expresar su satisfacción con un profesor no valora más la competencia que la interacción docente en la labor del profesor.

Parece más bien que lo que el alumno refleja a través de estos cuestionarios es la manifestación de su satisfacción con el profesor pero de manera general. Se pierde

por lo tanto una magnífica oportunidad de utilizar encuestas con carácter formativo.

Creemos necesario la optimización y mejora de estos procedimientos de evaluación. Los

alumnos no deben ser la única fuente de información en la evaluación del profesorado ni las encuestas el único procedimiento de evaluación utilizado.

7. BIBLIOGRAFÍA

Aparicio, J. J., San Martín, R. y Tejedor, F. J. (1982). La enseñanza universitaria vista por los alumnos: Un estudio para la evaluación de los profesores en la enseñanza superior. Madrid: I.C.E. Universidad Autónoma.

Apodaka, P. y Rodríguez, M. (1999). La opinión de los alumnos en la evaluación de la calidad docente: Posibilidades, limitaciones y estructura dimensional de sus indicadores. En Vidal, J. (Coords.) (1999) Indicadores en la universidad: Información y decisiones. Madrid: Consejo de Universidades. MEC

ApodaKa, P. (2001). Calidad y evaluación de la educación superior: Situación actual y prospectiva. Revista de Investigación Educativa, 19: 2, 367-382.

ApodaKa, P. y Grad, H. (2002). Análisis dimensional de las opiniones de los alumnos universitarios sobre sus profesores: comparación entre técnicas paramétricas y no-paramétricas. Revista de Investigación educativa, 20:2.

Bosker, R. y Snijders, T. (1999). *Multilevel Analysis: An Introduction to basic and advanced multilevel modeling*. London: Sage.

Bryk, S. y Raudenbush, W. (1992). *Hierarchical Linear Models*. Ca: Sage.

Bruce, A. J. (1985, Abril). *A Comparison of Three Teaching Evaluation Instruments. Ponencia presentada en la Convención of the Southwestern Psychological Association*.

Cohen, P. A. (1980). *Effectiveness of student-rating feedback for improving college instruction: A meta-analysis of findings*. Research in Higher Education, 13, 312-341.

Cruse, D. (1987). *Student evaluations and the university professor*. Higher Education, 15:6, 723-737.

Feldman, K. A. (1996). *Identifying exemplary teaching: Using data from course and teacher evaluation*. New Directions for Teaching and Learning, 65 ,41-50.

Fernández Sánchez, J. y Mateo, M. (1992). *Student evaluation of university teaching quality: Analysis of questionnaire for a sample of university students in Spain*. Educational And Psychological Measurement, 52, 675-684.

Goldstein, H.(1986). *Multilevel mixed linear model analysis using iterative generalized least squares*. Biometrika , 73, 43-56.

Goldstein, H. (2003). *Multilevel Statistical Models*. Third Edition. London: Arnold.

González Valverde, P. y Grande Quejigo, F. (1999). Experiencia en la evaluación de la universidad. El caso del profesorado. Revista Interuniversitaria de Formación del Profesorado, 34, 61-67.

Hox, J.J. (2002). *Multilevel Analysis, Techniques and applications*. London:Lawrence Erlbaum .

Leyland, A.H. y Goldstein, H. (2001). *Multilevel Modelling of health statistics*. New York: Wiley.

Marhs, W. (1997) *Making Students' Evaluations of Teaching Effectiveness Effective: The Critical Issues of Validity, Bias, and Utility*; American Psychologist, Vol. 52,