

**Universidad de Costa Rica.
Vicerrectoría de Investigación**

Instituto de Investigación En Educación

**INFORME FINAL DEL PROYECTO DE INVESTIGACION
UN MODELO CAUSAL EXPLICATORIO DE LOS
FACTORES UTILIZADOS EN LA EVALUACIÓN
DOCENTE UNIVERSITARIA**

Nº 720-A7-054

Raziel Acevedo Álvarez

Año 2008

INDICE

RESUMEN EJECUTIVO.....	4
1. ANTECEDENTES.....	5
I. Introducción.....	5
1. Código del proyecto:	5
2. Título del proyecto:.....	5
3. Unidad de pertenencia:	5
4. Unidad de adscripción.....	5
5. Programa al que pertenece.....	5
6. Nombre del investigador y carga académica.....	5
7. Duración del proyecto.....	5
II. Antecedentes del proceso investigativo.....	6
III. Planeamiento Del Problema.....	7
IV. Objetivos Del Proyecto.....	7
4.1. Objetivo General.....	7
4.2. Objetivos específicos	7
2. MARCO TEÓRICO REFERENCIAL	8
3. METODOLOGÍA.....	12
4. RESULTADOS.....	15
4.1. Análisis del modelo estructural (segundo objetivo).....	15
4.2. Elaboración de artículo (tercer objetivo)	21
4.3. Conclusiones	22
4.4. Dificultades	23
4.5. Recomendaciones	24
4.6. Publicaciones.....	25
5. IMPACTO.....	25
6. BENEFICIO RECIBIDO PARA LA UNIVERSIDAD DE COSTA RICA...25	
7. BIBLIOGRAFÍA CITADA Y CONSULTADA.....	26
.....	42
ANEXOS.....	43
ANEXO N° 1 Síntesis: La Evaluación Docente.....	43
ANEXO N° 2 Factores De Evaluación Docente.....	51
ANEXO N° 3 Artículo Concluido.....	60
INDAGACIONES TEÓRICAS ACERCA DE LA FIABILIDAD Y VALIDEZ EN LA EVALUACIÓN DOCENTE.....	60

Perfil académico y profesional de los autores.....	60
Dirección de los autores.....	60
Resumen	60
Abstract:	60
INTRODUCCIÓN:	61
1. Los cuestionarios en la medida de la competencia docente:.....	62
2.- La fiabilidad, un paso inicial de estudio.....	62
2.1. La unanimidad.....	64
2.2. La estabilidad.....	65
3. La validez en los cuestionarios una operación complicada.....	67
3.1. Los años 70 y la nota final	68
3.2. De los 80 hasta el presente el uso de otras técnicas.....	68
3.3. Nuevas perspectivas de estudio y análisis.....	71
3.4. Fuentes de sesgo.....	74
3.5 Correlaciones positivas entre las dimensiones e indicadores de competencia docente	76
4. DISCUSIÓN	77
5. A MANERA DE CONCLUSIÓN Y CIERRE.....	80
6. REFERENCIAS.....	81

RESUMEN EJECUTIVO

Un Modelo Causal Explicatorio De Los Factores Utilizados En La Evaluación Docente Universitaria

Dr. Raziel Acevedo Álvarez

El cuestionario utilizado para conocer la opinión de los (las) estudiantes respecto de la acción docente, es comúnmente empleado de acuerdo con diferentes intereses de los involucrados, llámese administradores (as), profesores (as) y estudiantes (as). Por ejemplo: 1) la administración los tienen en cuenta, como parte del proceso de toma de decisiones relacionadas con la selección de su personal, los incentivos de promoción económicos o académicos y con la permanencia de sus docentes dentro del sistema universitario; 2) a los instructores (as), les suministra información actualizada acerca de la visión que tienen sus estudiantes de la forma y estructura de sus cursos, de los elementos a mejorar y de la calidad de su enseñanza; 3) a los (as) estudiantes les proporciona herramientas que les permite seleccionar los cursos y las personas de su interés; y 4) a los investigadores (as) les es vital, para indagar sobre la eficacia de la enseñanza y el aprendizaje en la universidad. En fin, los resultados aportados por los cuestionarios han sido utilizados para una gran variedad de fines, por tanto es necesario tener la certeza de que estos instrumentos son totalmente fiables y válidos.

Pese a la importancia de los cuestionarios, según Marsh (2003) muy pocos de ellos han sido estudiados profundamente en cuanto a su fiabilidad y validez, es más cuentan con una insuficiente estructura teórica; lo cual incrementa la incertidumbre y el rechazo por parte de los docentes universitarios (as), sobre la calidad, la rigurosidad del instrumento y la relevancia de los datos aportados. En este sentido, la medida de la competencia docente deja una gran cantidad de interrogantes iniciales relacionadas con la fiabilidad y validez.

El Centro de Evaluación Académica ha identificado seis factores sobre los cuales se erige la estructura del cuestionario de evaluación docente. Estos son: evaluación y temática, (factores exógenos), y tres factores mediadores endógenas, trato, reglamento y didáctica, las cuales influyen directamente sobre el criterio u opinión general del estudiante. Cada uno de estos ha sido plenamente identificado en la literatura internacional sobre el tema, eso quiere decir que nuestra universidad utiliza factores que se han utilizado en otras universidades.

Estos factores constituyen un modelo teórico que puede ser representado gráficamente para su estimación empírica. Este modelo propone que la evaluación y la temática, covarian conjuntamente y que influyen directamente sobre el trato, la reglamentación y la didáctica. A su vez, estos tres factores, ejercen influencia directa sobre el criterio general que el estudiantes acerca de la labor de su profesor (a).

1. ANTECEDENTES

I. INTRODUCCIÓN

1. Código del proyecto:

N° 720-A7-054

2. Título del proyecto:

Un Modelo Causal Explicatorio De Los Factores Utilizados En La Evaluación Docente Universitaria

3. Unidad de pertenencia:

Sede de Guanacaste

4. Unidad de adscripción

Instituto de Investigaciones Educativas

5. Programa al que pertenece

Programa de Investigación en el cambio, desarrollo y gestión de la Educación Superior No. 724-A5-900

6. Nombre del investigador y carga académica

Dr. Raziel Acevedo Álvarez T.C.; más un ¼ de tiempo adicional para el desarrollo del proyecto.

7. Duración del proyecto

Originalmente se planteó para llevarlo a cabo durante el año 2007. Sin embargo, por diversas razones, solicitamos ampliación para el 2008. De esa forma se introdujo el objetivo tres, que en principio no existía, a fin de dar mayor uso al material de lectura utilizado durante la investigación. Por medio de ese objetivo se desarrolló el artículo científico.

II. ANTECEDENTES DEL PROCESO INVESTIGATIVO

El tema sobre educación superior que mayor número de debates, literatura e investigación ha generado durante los últimos años, es la evaluación del profesorado universitario. Su amplio desarrollo se puede justificar y caracterizar fundamentalmente, porque la eficacia docente es un componente básico en el proceso de enseñanza-aprendizaje y, en consecuencia, la calidad de la universidad y su reputación, se sustenta en la calidad de sus profesores.

En relación con el tema se ha generado un gran debate, que en determinados contextos intentan relacionar los resultados de estas evaluaciones con diferentes variables, con factores de sesgo o con factores de eficacia docente (Al respecto se puede ver: Craton y Smith, 1990; March, 1997, 2000; Cohen, 1986; Aleamori, 1999; Abrami y otros, 1997; McKeachie, 1990; Feldman, 1997).

Los estudios relacionados con el sesgo y representados entre otros por Centra (1993), Aleamori (1999), Feldman (1997), Marsh y Roche (1997) y Fernández (1998), ponen de manifiesto las variables de personalidad o comportamiento del profesor pueden verse afectadas por el tamaño de la clase, la obligatoriedad, el sexo o la disciplina académica, entre otras.

Una segunda línea de estudios de mayor aceptación y desarrollo, se ha dedicado a buscar dimensiones o factores que caractericen al "buen profesor" y que puedan ser utilizados como predictores de evaluación docente universitaria. Acerca de este tema es posible encontrar en la literatura una cantidad considerable de propuestas de diferentes contextos, entre las que destacan Marsh (1984), Centra (1990), Abrami (1990) y Feldman (1976^a y 1976^b), y otros. En este caso, nuestra investigación se ajusta a esta línea, pues pretende determinar matemáticamente los predictores fundamentales para la evaluación docente en nuestra universidad. Con ello buscar un modelo de ecuaciones estructurales que explique el grado de asociación de los factores de evaluación utilizados por el C.E.A.

Alrededor de los factores de evaluación, existe de una pequeña concordancia entre los investigadores y administrativos, sobre algunas dimensiones utilizadas en los instrumentos desarrollados, pues estos se elaboran alrededor de factores relacionados con organización, evaluación e interacción, entre otras. Sin embargo, aunque existe cierta relación entre los factores, todavía no se encuentran referencias claras sobre qué dimensiones son las predictoras más importantes de la competencia docente y cuántas son necesarias para obtener una visión general del profesor. Abrami, d'Apollonia y Cohen (1990) señalan que los predictores utilizados para la medida de la competencia docente varían dependiendo de los aspectos de la docencia que estén siendo objeto de estudio, por esa razón no existen investigaciones que identifiquen concretamente tales elementos.

Ahora bien, si bien es cierto que existe una voluminosa literatura producida por este campo, según Marsh y Roche (1997) son pocas las investigaciones que han sido evaluadas profundamente y con la dedicación necesaria para atender los aspectos relacionados la

validez de constructo, la cual es comprobada por medio de los modelos de ecuaciones estructurales. Esto ha creado una fuerte desconfianza y resistencia por parte de los profesores; muchos tienen miedo a la evaluación porque no confían en los instrumentos o por desconocimiento de sus formas y objetivos, o por la falta de credibilidad de las instancias evaluadoras, o por planteamiento apresurado acompañado de un estado de desinformación total o por la escasa validez de los instrumentos empleados para tal fin.

Por esa razón, es necesario desarrollar este proyecto de investigación, para continuar con la búsqueda de un modelo estructural, que explique con pertinencia y ponderación la validez del instrumento desarrollado por el C.E.A. En el pasado pudimos analizar un modelo básico, pero es necesario e importante continuar desarrollando modelos alternativos que expliquen con mayor profundidad la estructura planteada por nuestra universidad, a fin de brindar a los usuarios la confianza total en dicho instrumento de medida. Al mismo tiempo, tales argumentos permitirán en el futuro elaborar instrumentos de medida más precisos y confiables.

III. PLANEAMIENTO DEL PROBLEMA

El estudio de la competencia docente, desde la perspectiva de los estudiantes, plantea varias interrogantes de estudio, que pueden ser abordadas desde:

¿Qué factores se han asociado positivamente con la docencia universitaria?

¿Cuáles son los factores que a nivel internacional se han empleado con mayor frecuencia?

¿Existe un constructo en el instrumento de evaluación propuesto por el C.E.A.?

Para poder atender estas y otras interrogantes en este proceso de estudio, se diseñan los siguientes objetivos.

IV. OBJETIVOS DEL PROYECTO

4.1. Objetivo General

Identificar y ponderar los factores utilizados como criterios de medida de la competencia docente universitaria, a partir de la percepción de los estudiantes, con el fin de validar o rechazar el constructo propuesto por el Centro de Evaluación Académica de la U.C.R.

4.2. Objetivos específicos

Construir y estimar el modelo causal que explique la mayor cantidad de varianza entre los factores del instrumento de evaluación propuesto por el C.E.A, para aplicar un análisis factorial confirmatorio que asocie matemáticamente las relaciones causales entre estos.

Observar los índices de ajuste del modelo que muestren las bondades o problemas del constructo, para determinar la validez del instrumento del C.E.A.

Elaborar un artículo con el material de análisis y registro, para por un lado mostrar el peso del constructo propuesto por la U.C.R. y por otro, exponer la fuerza de los modelos de ecuaciones estructurales como herramienta de investigación cuantitativa.

2. MARCO TEÓRICO REFERENCIAL

El estudio de la evaluación de la competencia docente, por parte de los estudiantes, ha sido una preocupación permanente en el campo de la investigación educativa universitaria. Cada año se publican diferentes perspectivas teóricas, que muestran los dos grandes campos del saber: a) los factores asociados positivamente con la competencia docente y b) los factores que sesgan este tipo de actividad.

Aunque a lo largo de los años aparecen diversas propuestas y estudios, existen pocas variantes substanciales que muestren una innovación en la docencia universitaria y muchos estudios presentan diversas dimensiones o factores, que resultan ser los mismos propuestos por otros autores, pero con diferente nombre.

La búsqueda de los factores asociados positivamente a la docencia universitaria es clave, en el proceso de desarrollo y mejora de las universidades, porque con herramientas construidas adecuadamente se pueden llegar a ejercer cambios a nivel institucional y particular, que puedan incidir directamente en la mejora de la docencia universitaria.

A continuación se listan las principales publicaciones que fueron revisadas y que se utilizaron como estímulo para la discusión y propuesta de un modelo de ecuaciones estructurales.

Abalde, E.; De Salvador, X; González Carbanach, R. Y Muñoz C.; J.M. (1995). Análisis de la Evaluación de la Docencia Universitaria por los (as) alumnos (as) en la Universidad de la Coruña (1993-1994). **En Estudios de Investigación Educativa en Intervención Pedagógica**, pp. (289-292) Valencia: AIDIPE.

Abrami, P.S.; D'apollonia, S. & Rosenfield. (1997). .The Dimensionality Of Student Ratings of Instruction. What We Know and What We do Not.. R.P. PRRY & J.C. SMART (Eds.) **In Effective Teaching in Higher Education. Research and Practice**, pp.321-367. New York. Agathon Press.

Acevedo Alvarez, R. & Rodríguez, N. M. (2006). Factores de sesgo asociados a la validez de la evaluación docente universitaria: un modelo jerárquico lineal. **Archivos Analíticos de Políticas Educativas**, **14** (34). Recuperado [27 marzo 2007] de <http://epaa.asu.edu/epaa/v14n34/>

Acevedo, R. y Fernández, M. J. (2004): "La percepción de los estudiantes universitarios en la medida de la competencia docente: validación de una escala". En **Revista de Educación**. N° 28 (2): 145-166. San José: Editorial de la Universidad de Costa Rica.

Aleamori, L.M. (1999). Student Rating myths versus research facts from 1924 to 1998.. **Journal of Personnel Evaluation in Education**, **13** (2), pp. 153-166.

Braskamp, L.A. & Ory, J.C. (1994). Assessing Faculty Work. hancing Individual and Institutional Performance. San Francisco. Jossey-Bass.

Braskamp, L.A.; Ory, J.C. & Pieper, D.M. (1981). Student Written Comments. Dimensions of Instructional Quality. **Journal of Educational Psychology**, **73** (1), pp. 75-70.

Centra, J.A. (1993). Reflective Faculty Evaluation. hancing Teaching and Determining Faculty Effectiveness. San Francisco: Jossey-Bass.

Coffey, M. & Gibbs, G. (2001). The Evaluation of the Student Evaluation of Educational Quality Questionnaire (SEEG) in U.K. Higher Education. **Assessment & Evaluation in Higher Education**, **Vol 26**, (1), pp. 89-93.

Craton, P. y Smith, R. A. (1990). Reconsidering the Unit of Analysis. A Model of Student Ratings of Instruction. **Journal of Educational Psychology**, **82** (2) pp. 207-212.

D'apollonia, S. & Abrami, P. C. (1997). Navigating Student Ratings of Instruction. **American Psychologist**, **51** (11), pp. 1198-1208

Feldman, K. A. (1997). Identifying Exemplary Teachers and Teaching. Evidence from Student Ratings. R. P. En PERRY & J. C. SMART (eds.), **Effective teaching in Higher Education. Research and Practice**, . (pp. 368-395). Bronx, N. Y: Agathon.

Fernández, J. ; Mateo, M. & Muñíz, J. (1998). Is There Relationship Between Class Size and Student Ratings of Teacher Quality?. **Educational And Psychological Measurement**, **58** (August), 596-604

García Ramos, J. M. (1999^a). Análisis multirasgo- multimétodo en la validación de instrumentos de medida para la evaluación de la calidad docente en instituciones universitarias. **Revista Española de Pedagogía**, **214**, pp. 417-444.

García Ramos, J. M. (1997). Valoración De La Competencia Docente Del Profesor Universitario. Una Aproximación Empírica. **Revista Complutense De Educación**, **8** (2), pp. 81-108

Goddard, R; Hoy, W & Woolfolk, A. (2001). Collective Teacher Efficacy. Its Meaning, Measure, and Impact on Student Achievement. **Collective Teacher Efficacy**. Manuscrito sometido para publicación. 1-40.

Greenwald, A. G. Y Gillmore, G. M. (1997). Grading Leniency is a Removable Contaminant of Student Ratings. **American Psychologist** , **51** (11), 1209-1217.

Jornet, J. M. ; González Such, J. Y Pérez Carbonell, A. (1995) . Evaluación De la Actividad Universitaria. G. Quintás (eds.), En **Reforma y Evaluación de la Universidad**. Valencia: Servei de Publicacions de la Universitat de Valencia. pp. 189-244.

Marsh, H. & Roche, L. R. (1997). Making Students' Evaluation of Teaching Effectiveness Effective. The Critical Issues of Validity, Bias, and Utility. **American Psychologist**, **Vol. 52**, 11, 1187-1197.

Marsh, H. y Bailey, M. (1993). Multidimensional Students' Evaluations of Teaching Effectiveness. **Journal of Higher Education**, **64** (1), pp. 1-18.

Marsh, H. A. y Overall, J. U. (1979b). Validity of student's evaluation of teaching. A comparison with self evaluations by teaching assistants, undergraduate faculty, and graduate faculty. Paper presented Annual Meeting of the American Educational Research Association, San Francisco.

Marsh, H. W. (1977). The validity of students' evaluations of instructors independently nominated as best and worst teacher by graduating senior. **American Educational Research Journal**, **14**, pp. 441-447.

Marsh, H. W. (1980). The influence of student, course, and instructor characteristics on evaluations of university teaching. **American Educational Research Journal**, **17**, 219-237.

Marsh, H. W. (2001). Distinguishing between good (useful) and bad workloads on Students' Evaluations of teaching. **American Educational Research**, **38** (1), pp. 183-212.

Marsh, H. W. y Dunkin, M. J. (1992). Students' Evaluation of University Teaching. A Multidimensional Perspective. J. SMART (ed.) **Higher Education. Handbook of Theory and Research**. (pp. 143-223) New York: Agathon.

Marsh, H. W. y Overall, J. U. (1981). The relative influence of course level, course type, and instructor on students' evaluations of college teaching. **American Educational Research**, **18**, pp. 103-112.

Marsh, H. W. y Roche, L. A. (2000). Effects of grading leniency and low workload on students' evaluation of teaching. popular myth, bias, validity or innocent bystanders?. **Journal of Educational Psychology**, **92** (1), pp. 202-228.

Marsh, H. W. ; Hau, K. T. ; Chung, C. M. & Siu, T. L. P. (1997). Students' s Evaluations of University Teaching. Chinese Version of The Student's Evaluations of Educational Quality Instrument. **Journal of Educational Psychology**, **89** (3), pp. 568-572.

McKeachie, W. J. (1997). Student Ratings. The Validity of Use. **American Psychologist**, **52** (11), pp. 1218-1225.

Rindermann, H. & Schofield, N. (2001). Generalizability of Multidimensional Student Rating of University Instruction Across Courses and Teacher. **Research in Higher Education**. Vol. **42**, 4, pp. 377-400.

Salvador, L. (1990). Los Docentes Universitarios Exitosos Desde La Perspectiva del Alumno. Su Caracterización Psicopedagógica. Tesis Doctoral. Universidad de Salamanca. España.

Saroyan, A. & Amundsen, Ch. (2001). Evaluating University teaching. Time to Take Stock. **Assessment & Evaluation in Higher Education**, Vol **26**, (4), pp. 341-353

Timpson, W. W. Y Andrew, D. (1997). Rethinking student evaluation and the improvement of teaching. instrument for change at the University of Queensland. **Studies in Higher Education**, **22** (1), pp. 55-65.

Ting, K. F. (2001). A Multilevel Perspective On Student Ratings of Instruction. Lessons From the Chinese Experience. **Research in Higher Education**. Vol. 41, 5, pp. 637-653.

Villa, A. Y Morales, P. (1993). **La Evaluación Del Profesor. Una Visión De Los Principales Problemas Y Enfoques De Diversos Contextos**. Vitoria: Departamento de Educación, Universidades e Investigación. Gobierno Vasco.

Wachtel, H. K. (1998). Student Evaluation of College Teaching Effectiveness. A Brief Review. **Assessment & Evaluation In Higher Education**, 23 (June), 191-211.

Wigington, H. ; Tollefson, N. Y Rodríguez, E. (1989). Student's ratings of instructor visited. Interactions among class and instructor variables. **Research in Higher Education**, 30 (3), pp. 331-334.

Wilkerson, D. ; Rogers, M. A. Y Maughan, R. (2000). Validation of Student, Principal, and Self Ratings in 360 Feedback for Teacher Evaluation. **Journal of Personnel Evaluation in Education**, 14 (2), pp. 179-192

Worthington, A. C. (2002). The impact of student perception and characteristics on teaching evaluation. a case study in finance education. **Assessment & Evaluation in Higher Education**, 27 (1), pp. 49-64.

Vale la pena anotar que esta es una pequeña muestra de la gran cantidad de trabajos existentes a nivel de lengua inglesa y lengua castellana. A todas luces, se muestran pocos trabajos a nivel de Centroamérica. Por lo anterior, es importante comenzar a hacer conciencia en todos los investigadores que trabajan esta temática y en los personeros del C.E.A., sobre la importancia de construir un cuerpo teórico de referencia por medio de publicaciones, ya sea de los informes de trabajo u otros similares, de manera tal que permita mostrar a otros investigadores o interesados, la estructura utilizada en Centroamérica.

3. METODOLOGÍA

Inicialmente el proyecto se dividía en dos componentes: lectura sobre el tema y la estructuración, estimación y análisis del modelo de ecuaciones estructurales a seguir. No obstante, al contar con una gran cantidad de material escrito y ordenado en nuestra base de datos bibliográfica, se pensó en solicitar una prórroga, para incluir un tercer objetivo que permitiera la elaboración de un artículo científico con todo el material de la investigación. De esa forma, para el año 2008, con la prórroga señalada se procedió a elaborar un artículo

sobre la temática de la investigación, fundamentado directamente en los estudios de fiabilidad y validez de los instrumentos de evaluación docente universitaria. Los dos objetivos iniciales, se cumplieron en el 2007 y se anotaron en el informe parcial de investigación.

La investigación propuesta se estructura alrededor del paradigma cuantitativo, por cuanto, el investigador no se encuentra inmerso en el problema de estudio, sino está independiente de él, es más, nunca tuvo contacto con la población muestra, pues el C.E.A. se encargó de construir y aplicar el instrumento de medida de la competencia docente en varias facultades y Sedes de nuestro país.

Este instrumento correspondía a un cuestionario cerrado de 53 ítems, otro elemento del paradigma utilizado, que se aplicó por los personeros del C.E.A. a una muestra de 14400 estudiantes durante el 2006, en diferentes Sedes y Facultades de nuestra universidad.

Con los datos a mano, organizamos e integramos cada una de las bases de datos facilitadas por el C.E.A., para ajustar una estructura propia de los modelos de ecuaciones estructurales, una metodología utilizada para el análisis estadístico en la comprobación del análisis factorial confirmatorio de los factores propuestos por el C.E.A.

Los modelos de ecuaciones estructurales tienen su fundamento en los estudios de regresión, que observan el grado de asociación entre las variables. Indudablemente, estos son más profundos y permiten observar el peso de una variable sobre otra, su medida de error y sobre todo, su estimación facilita diversos índices que señalan el ajuste o desajuste del modelo.

El modelo se plantea en el paquete informático, realizando un diagrama con variables exógenas y endógenas. A cada variable le corresponde uno o varios ítems medibles los cuales aportan información sobre el factor a estudiar.

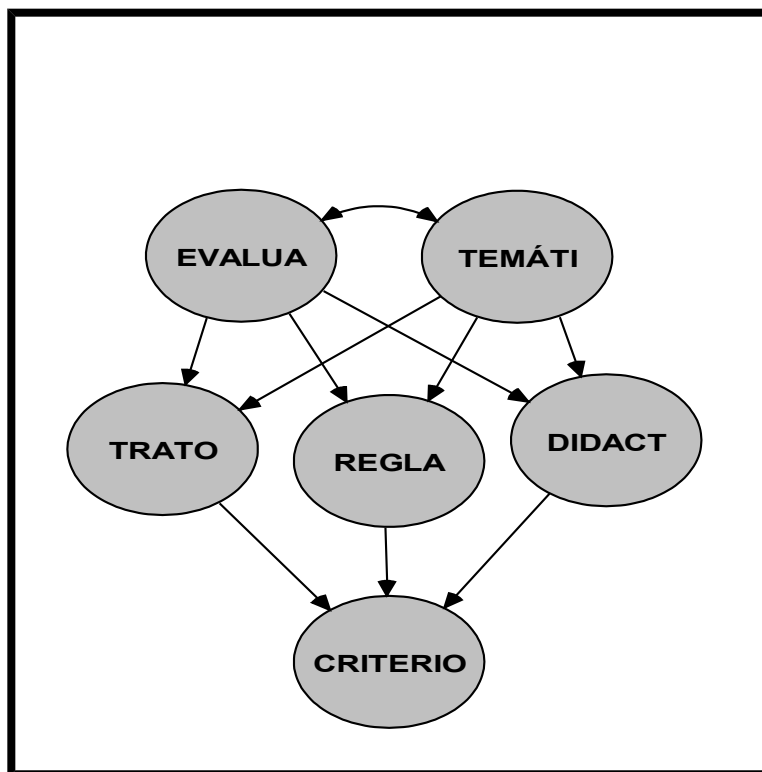
Con los datos a mano, organizamos e integramos cada una de las bases de datos facilitadas por el C.E.A., para ajustar una estructura propia de los modelos de ecuaciones estructurales para ingresar esos datos en alguno de los paquetes estadísticos utilizados por los modelos de ecuaciones estructurales. Este trabajo fue muy minucioso y le dedicamos una gran cantidad de horas a cada una de las siete bases, subdivididas en 24 Facultades o Escuelas. El trabajo de organización se logró en un 100%

Integrada la base de datos, solicitamos una cita al C.E.A. para proponer un modelo, que consiste en especificar una estructura gráfica que determine las relaciones causales entre los factores. Para generar el modelo, este investigador se fundamentó en una publicación del C.E.A. a cargo de M.Sc. Lorena Kirkuk.

Elaborado el modelo se cumple al 100% el objetivo N° 1, identificando los parámetros asociados a los constructos del instrumento. El modelo tiene dos variables exógenas, evaluación y temática, que se encuentran asociadas entre sí e influyen directamente sobre tres variables mediadoras endógenas, trato, reglamento y didáctica, las

cuales influyen directamente sobre el criterio. El modelo se presenta a continuación, como se elaboró en el paquete informático AMOS

MODELO FACTORIAL CONFIRMATORIO



Como lo propone el modelo, se cuenta con dos variables exógenas, evaluación y temática, que se asocian entre sí e influyen directamente sobre tres variables mediadoras endógenas, trato, reglamento y didáctica, las cuales tienen relación directa sobre el criterio, o sea, la valoración global que brinda el estudiante al docente. Este modelo cuenta con siete factores, seis de ellos se encuentran asociados al factor criterio del estudiante.

El objetivo dos, tiene que ver directamente con los resultados producidos por la estimación del modelo, lo cual anotaremos en el apartado siguiente de resultados. De la misma forma, el objetivo tres, lo adjuntamos en ese mismo capítulo. Aunque su metodología fue diferente, pues trató de ordenar la información recopilada durante el proceso de lectura sobre el tema. Con ese ordenamiento, se procedió a organizar un artículo científico.

En síntesis podemos mencionar que esta investigación cuantitativa, es transversal correlacional, que eventualmente podría convertirse en longitudinal si cada año o cada dos años, el C.E.A. u otros investigadores continuaran elaborando o mejorando los modelos

teóricos propuestos por la entidad, de esa forma se contaría con información sobre el instrumento aplicado en otros años.

4. RESULTADOS

Con el material que a continuación se desarrolla, fuimos invitados por Universidad Nacional De Ingeniería de Nicaragua, para impartir charlas sobre la evaluación docente, para apoyar las gestiones que realiza esta institución sobre nuestro tema de investigación. En las tres jornadas realizadas, participaron alrededor de 20 personas de los distintos departamentos que tienen relación con la docencia universitaria. Al finalizar, fuimos felicitados por los vicerrectores de la institución e invitados a construir con ellos sus instrumentos de medida.

Adicional, durante el Banco Mundial, nos invitó a Nicaragua para apoyar el departamento de estadística y educación, en su proceso de análisis de datos de educación y evaluación docente, debido a las recomendaciones brindadas por la Universidad Nacional De Ingeniería. Con los personeros del M.E.P. estuvimos trabajando cuatro días sobre los diferentes temas que se reúnen alrededor de la docencia.

Concientes de la problemática de los métodos de investigación cuantitativas, ya sea por el poco uso, por desconocimiento y o por la falta de herramientas en los lugares de trabajo de los investigadores, planificamos un curso de extensión docente sobre metodología de análisis de datos en el SPSS, en el cual participaron investigadores, estudiantes y profesionales extranjeros. El curso pretende ir generando posibilidades de desarrollo de un taller permanente de Acción Social para el 2009.

4.1. Análisis del modelo estructural (segundo objetivo)

Vale la pena anotar, que aquí se presentan los resultados más relevantes del estudio desde el punto de vista sustantivo y académico. La información y las bases analizadas son muy grandes, por ello, a razón de presentar un informe sintético hemos decidido suprimir mucha información no necesaria y mantener solamente aquella que realmente es vital e importante para conseguir los resultados del estudio.

Conseguimos interactuar con los investigadores del C.E.A. para adquirir una base de datos reales, sobre algún instrumento reciente aplicado a los profesores de nuestra institución. Este paso se logro en un 100% pues los encargados del C.E.A. gustosamente y ágilmente facilitaron una gran cantidad de información de aplicaciones recientes en las diferentes facultades de la U.C.R. sobre el nuevo instrumento de evaluación del docente universitario.

Solicitamos nuevamente al Centro de Informática un paquete estadístico para el manejo de los modelos de ecuaciones estructurales, a fin de implementar las complejas fórmulas matemáticas, que posibiliten el estudio de la estructura propuesta por el C.E.A. Lo que en principio parecía el paso más simple, resultó ser el más difícil dado que este centro no cuenta con ninguno de los paquetes como: AMOS, LISREL o EQS y desde el 2004

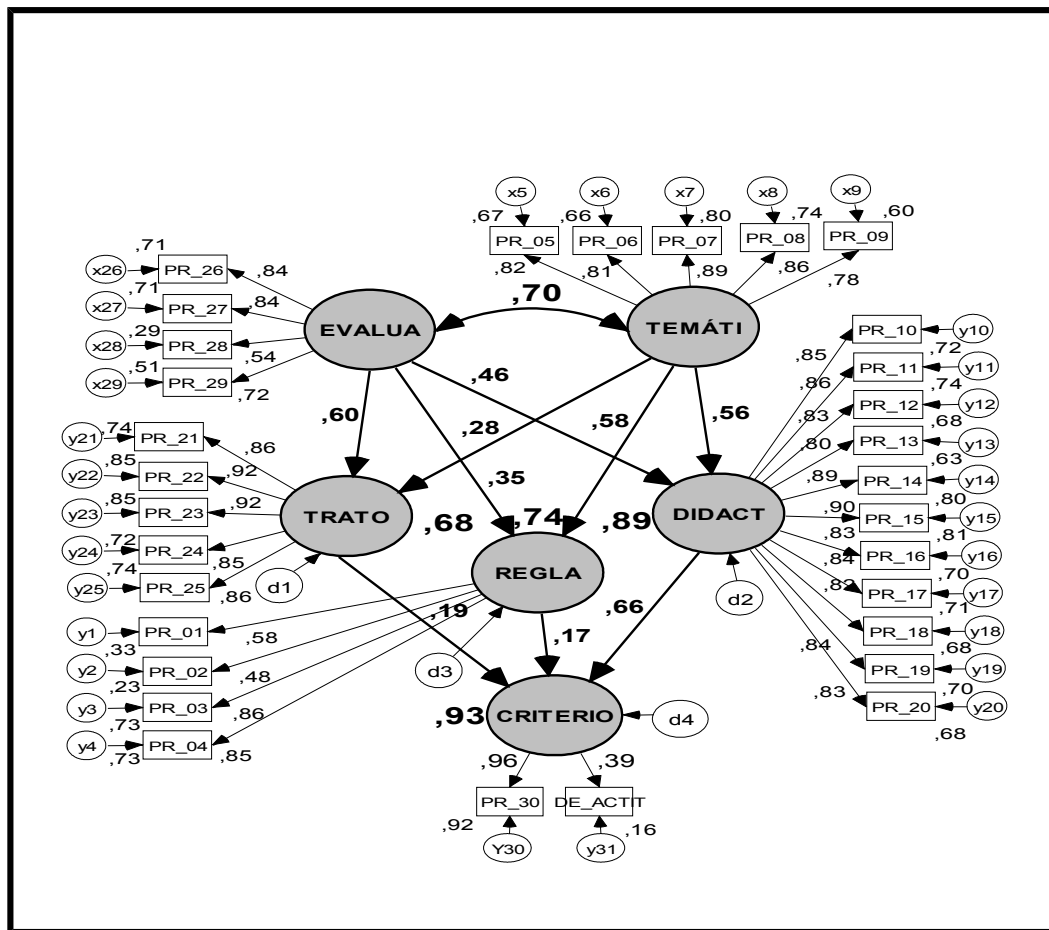
estamos tratando de que se adquieran. Por ello utilizamos nuestros paquetes personales para desarrollar el proyecto.

Elaborado el modelo de competencia docente por medio del la propuesta de análisis factorial exploratorio de Kikut (2003), cumplido el objetivo 1, nos dimos a la tarea de proceder con el análisis, el cual correspondía al segundo objetivo:

“Observar los índices de ajuste del modelo que muestren las bondades o problemas del constructo, para determinar la validez del instrumento del C.E.A.”

Antes de iniciar la estimación del modelo; se ajustaron a uno los parámetros de los siguientes ítems: Reglamento ítem N°1; Temática ítem N°8; Didáctica ítem N°17; Trato ítem N°22 y el ítem N°28 del factor Evaluación. Esto debido en primer, lugar a las especificidades del programa estadístico y en segundo, por la alta saturación de la carga factorial explicada por cada uno de los ítem mencionados en el texto de Kikut (2003) p. 17.

Estimado el modelo, se obtienen los resultados mostrados en el siguiente gráfico.



Una aproximación visual de los resultados anteriores nos permite determinar la existencia de relaciones causales entre los factores y sus correspondientes ítems. Lo cual demuestra la existencia de un modelo teórico dominado por dos factores exógenos y cuatro endógenos.

A primera vista es evidente que la mayoría de los indicadores de cada uno de los factores son buenos y excelentes sus r cuadrados, destacando con esa información su alta relación con el factor que representa. Estos indicadores demuestran el peso de cada factor, como: TRATO con 0,68; REGLAMENTO con 0,74; DIDACTICA con 0,89 y CRITERIO con 0,93. Estos índices son muy altos y sugieren que los ítem que representan el factor se encuentran muy asociados con el mismo.

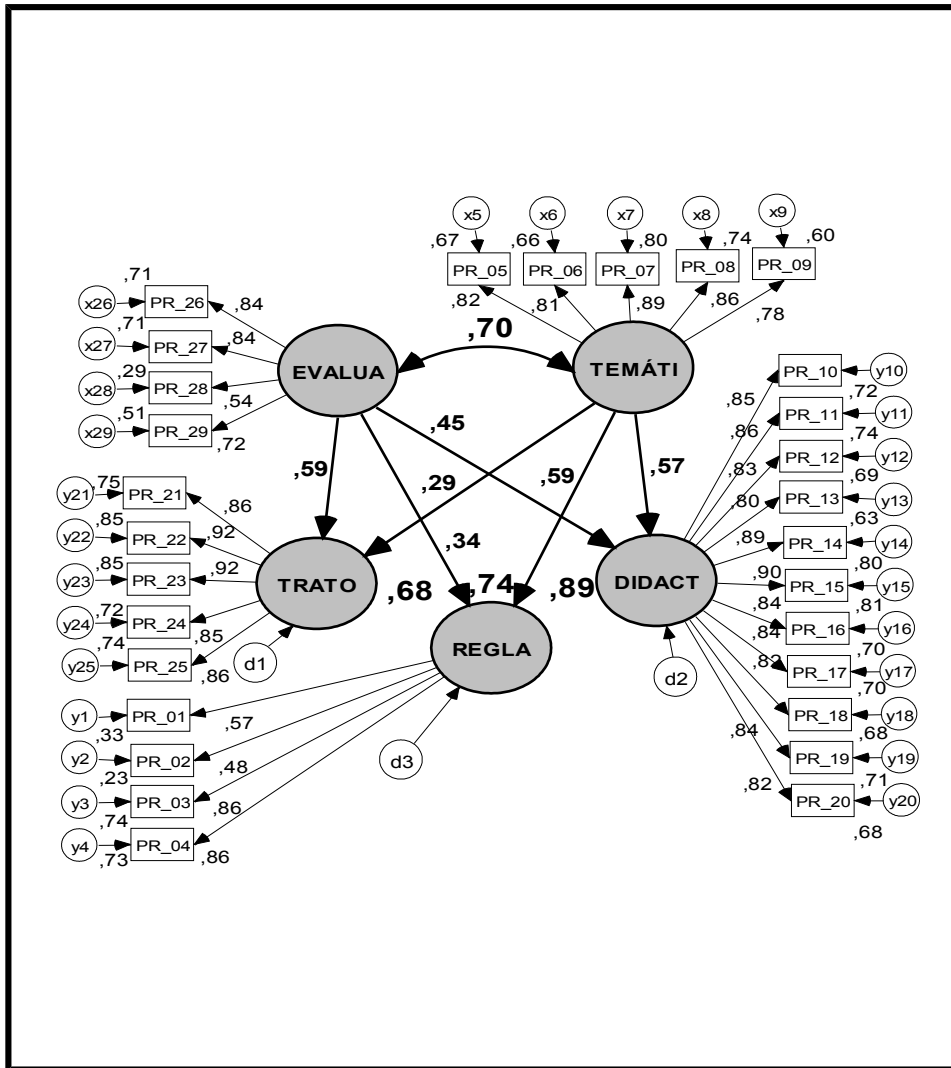
Por otra parte, el modelo propone una fuerte asociación entre las variables exógenas EVALUACIÓN y TEMÁTICA, 0.70, un índice muy alto para este tipo de modelos. Además se observa una bajo r en los ítems: 31 (0.39); 02 (0.48); 28 (0.54).

Sin embargo, al revisar los índices de ajuste algunos nos dejan ciertas dudas, especialmente el RMSEA 0,843 y el NFI 0,9061 (ver anexos). Esto podría deberse a que algunos ítem se encuentran poco representados o a la cantidad de relaciones causales explicadas en el modelo.

Debido a lo anterior, consideramos necesario reespecificar el modelo, partiendo de la eliminación de la variable CRITERIO, pues la teoría sobre la evaluación docente señala su poca importancia en este tipo de modelos, dado que este factor solamente reúne la opinión general que tiene el estudiante de la actividad de su profesor.

Los ítems con bajo r cuadrado observados en el modelo anterior permanecerán dentro de su respectivo factor, pues no queremos de momento, modificar sin permiso u autorización el instrumento elaborado por el C.E.A,

Al reestimar nuevamente el modelo, se obtienen los siguientes resultados.



La reespecificación del modelo modifica ligeramente el peso y la relación de algunos factores, por ejemplo: la influencia de TEMÁTICA sobre DIDÁCTICA, REGLAMENTO y TRATO, se incrementa en 0.1. Igual sucede con la influencia del factor EVALUACIÓN sobre REGLAMENTO y DIDÁCTICA. El único parámetro que se reduce en 0,10 es el de EVALUACIÓN sobre TRATO.

El grado de asociación entre los dos predictores exógenos, EVALUACIÓN y TEMÁTICA, es alta (0,70), evidenciando una fuerte relación entre ambos, señalando que estos factores son muy importantes para medir la acción del docente universitario a partir de la visión de sus estudiantes.

El factor evaluación ejerce una fuerte influencia sobre TRATO con un 0,59. Pero su asociación es menor en los factores DIDÁCTICA (0,45) y REGLAMENTACIÓN (0,34, lo cual indica que los estudiantes perciben mejor los procedimientos de evaluación que utilizan sus profesores, cuando reciben un trato personal humanitario de parte de sus

profesores. El segundo TEMÁTICA, se encuentra muy asociado con DIDÁCTICA (0,57) y REGLAMENTACIÓN (0,59), pero con muy poco efecto sobre el TRATO (0,29). La información pone de manifiesto que el manejo de la temática de trabajo, como es lógico, permanece muy relacionada con los procedimientos didácticos del docente.

Analizando los efectos de los factores y el grado de asociación entre ellos, se evidencia una estructura sólida, fundamentada en cinco factores. Una vez desarrollado este análisis procedemos a revisar los índices de ajuste del modelo. Estos índices, tienen mayor profundidad y nos permiten tomar la decisión de aceptar o rechazar totalmente el modelo anterior.

El primero de ellos es el CMIN, que permite diferenciar nuestro modelo con un modelo nulo.

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	94	38627,2546	370	,0000	104,3980
Saturated model	464	,0000	0		
Independence model	29	411371,9113	435	,0000	945,6826

El valor mínimo de la discrepancia, es un índice relacionado con el chi-cuadrado. Generalmente, en la mayoría de los casos este valor no ajusta y nuestro caso no es la excepción. Sin embargo los investigadores proponen, poner más atención a otros índices, como el del valor mínimo de la función de discrepancia FMIN. Observado en la siguiente tabla

FMIN

Model	FMIN	F0	LO 90	HI 90
Default model	2,6550	2,6295	2,5854	2,6741
Saturated model	,0000	,0000	,0000	,0000
Independence model	28,2749	28,2450	28,1003	28,3902

El valor recomendado para que un FMIN sea bueno, debe ser un número menor a 3. Nuestro resultado es 2,65, lo cual es una buena medida de la parsimonia del modelo.

También, índices de PRATIO, PNFI y PCFI, permiten observar la parsimonia del modelo propuesto, pero en el ajuste de las medidas de evaluación del modelo. En la siguiente tabla se observan los resultados aportados por el programa. Un valor cercano a 1 es bueno.

Parsimony-Adjusted Measures

Model	PRATIO	PNFI	PCFI
Default model	,8506	,7707	,7714
Saturated model	,0000	,0000	,0000
Independence model	1,0000	,0000	,0000

Si observamos los índices de parsimoniosidad del modelo, ver tabla siguiente, todos los índices se encuentran por debajo de los criterios establecidos por los investigadores. Los tres valores destacado en la tabla oscilan entre 0,77 y 0,85, unos valores no muy alentadores para esta investigación

Ahora bien, una vez comparado el modelo con otro alternativo, se procede a observar los siguientes índices, los cuales deben permanecer cercanos a 1.0

Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	,9061	,8896	,9069	,8905	,9069
Saturated model	1,0000		1,0000		1,0000
Independence model	,0000	,0000	,0000	,0000	,0000

Observando los índices de ajuste del modelo, se encuentran todos con valores satisfactorios dentro de los índices de normalidad.

NCP

Model	NCP	LO 90	HI 90
Default model	38257,2546	37615,4040	38905,4010
Saturated model	,0000	,0000	,0000
Independence model	410936,9113	408830,6471	413049,4521

Para contar con una medida más profunda sobre los índices y el modelo, se debe de observar el *root mean Squire error of approximation*, o RMSEA

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	,0843	,0836	,0850	,0000
Independence model	,2548	,2542	,2555	,0000

Los valores observados en este índice no son satisfactorios, pues se encuentran sobre 0.08, el indicador máximo para conocer el error máximo de aproximación. En este sentido, el valor obtenido nos deja muchas dudas para aceptar el modelo.

Cada uno de los estadísticos siguientes forman una medida compuesta de bajo ajuste del modelo. Los modelos complicados reciben altos puntajes, estos índices se utilizan para comparar con otros modelos y no son una medida del mismo.

AIC

Model	AIC	BCC	BIC	CAIC
Default model	38815,2546	38815,6430		
Saturated model	928,0000	929,9175		
Independence model	411429,9113	411430,0312		

ECVI

Model	ECVI	LO 90	HI 90	MECVI
Default model	2,6679	2,6238	2,7124	2,6679
Saturated model	,0638	,0638	,0638	,0639
Independence model	28,2789	28,1341	28,4241	28,2789

HOELTER

Model	HOELTER	HOELTER
	.05	.01
Default model	157	165
Independence model	18	18

Analizando todos los índices de ajuste del modelo obtenidos por la estimación del mismo, tenemos muchas inquietudes en cuanto a los valores derivados por nuestra propuesta. Si bien, algunos resultados son buenos, otros lo son medianamente y el error es muy alto, lo cual indica una falta de estabilidad en la propuesta. Es necesario la estimación con otros valores o con modelos alternativos para saber cual es la propuesta teórica que se ajusta a la empírica.

4.2. Elaboración de artículo (tercer objetivo)

Finalmente, durante el año 2008 nos dimos a la tarea de realizar un artículo con el material obtenido por la amplia revisión teórica realizada durante el proceso de investigación. El objetivo nuestro estaba claro, pero el mayor interés se centraba sobre el tema de la fiabilidad y validez de los cuestionarios utilizados por los y las estudiantes, para evaluar la competencia docente universitaria. Ya que el estudio del instrumento por medio de los modelos de ecuaciones estructurales, no es otra cosa más que la búsqueda de la validez de constructo, o sea, el último componente o el más robusto, de todos los tipos de validez estudiados por los investigadores.

Sobre este tema existe una gran polémica que no podíamos dejar de lado, pues es una temática relevante y pertinente, debido a la importancia que tiene para las instituciones universitarias la calidad no solo de su personal docente, sino también, de los instrumentos de medida que apliquen a sus profesionales. Como nuestros resultados con el modelo, no son totalmente satisfactorios y además es información confidencial, pensamos continuar en la búsqueda de un modelo de mejor ajuste y que tenga menor error. No obstante, el material obtenido en la investigación iba más allá del modelo teórico del C.E.A, por esa razón,

preferimos escribir sobre el tema de fiabilidad y validez, que escribir sobre el modelo de prueba que no alcanzó nuestra satisfacción total. En el anexo N°3 se encuentra el artículo.

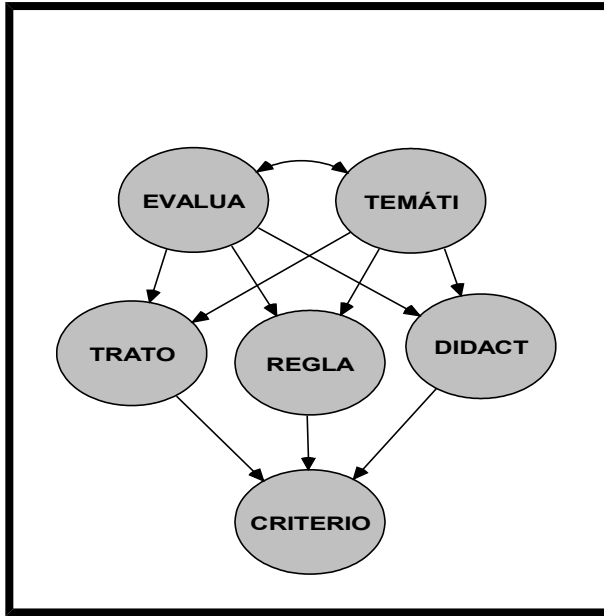
Para su elaboración se levantó una base bibliográfica con los trabajos más importantes sobre el tema. Una vez ordenada la información en la base de datos se digitaliza en FILE MAKER, el cual integra fichas divididas con la siguiente información: Autor, Bibliografía, Palabras claves, resumen y número de página. A continuación se procede a la construcción de un índice básico sobre los principales temas que tratará el artículo, en ello tardamos varios días. Este índice básico permite centralizar las ideas, a la hora de la redacción, y además, facilita la coherencia del mismo porque visualiza los apartados más importantes. El índice cambió varias veces para reordenar la estructura, a medida que se iba desarrollando los temas. Una vez concluida la redacción, se presentó a la revista electrónica del I.N.I.E. para su revisión.

4.3. Conclusiones

Fundamentado en el objetivo general de este trabajo que era “*Identificar y ponderar los factores utilizados como criterios de medida de la competencia docente universitaria, a partir de la percepción de los estudiantes, con el fin de validar o rechazar el constructo propuesto por el Centro de Evaluación Académica de la U.C.R.*”, podemos llegar a concluir que:

1.- En el análisis factorial exploratorio de Kikut (2003) y en el análisis factorial confirmatorio desarrollado por esta investigación, se han identificado seis factores sobre los cuales se erige la estructura del cuestionario de evaluación docente del C.E.A. Estos factores son: evaluación y temática, (factores exógenos), y tres factores mediadores endógenas, trato, reglamento y didáctica, las cuales influyen directamente sobre el criterio u opinión general del estudiante. Cada uno de estos ha sido plenamente identificado en la literatura internacional sobre el tema, eso quiere decir que nuestra universidad utiliza factores asociados con la buena competencia docente, los cuales se han utilizado en otras universidades.

2.- Estos factores constituyen un modelo teórico que puede ser representado gráficamente para su estimación empírica, de la siguiente forma (ob. 2). Este modelo propone que la evaluación y la temática, covarian conjuntamente y que influyen directamente sobre el trato, la reglamentación y la didáctica. A su vez, estos tres factores, ejercen influencia directa sobre el criterio general que el estudiantes acerca de la labor de su profesor (a).



3.- Los factores cuentan con un fuerte peso de representación de sus ítems, lo cual indica la existencia de una estructura factorial representada por cada uno de sus ítems. Sin embargo, al indagar en sus índices de ajuste incremental, parsimonia, RMSEA y otros indicadores de mayor precisión estadística, se observaron ajustes muy irregulares lo que expresa la inconsistencia del modelo, lo cual propone que hace falta mayor trabajo de este tipo con los modelos de ecuaciones estructurales, para llegar a obtener índices de ajuste totalmente satisfactorios y relevantes, los cuales permitirían contar con datos matemáticos totalmente exactos para demostrar la validez de la estructura propuesta por el C.E.A.

4.- De momento, no se puede estar totalmente seguros de la estructura, porque su ajuste no es del todo convincente, y tiene un error muy alto, lo cual deja muchas interrogantes de ajuste del modelo.

4.4. Dificultades

Como método de trabajo se fundamentó en la lectura de las bases de datos bibliográficas sobre el tema. Inicialmente hicimos búsqueda dentro del país y encontramos muy poco material sobre el tema, por tanto, ampliamos nuestro interés al ámbito internacional, donde logramos muchísimos documentos en idioma inglés y algunos trabajos realizados en España, principalmente.

Posteriormente reconstruyó una base de datos bibliográficos con FILE MAKER y con los datos empíricos aportados por las aplicaciones reunidas por el C.E.A. se analizaron utilizando la metodología de modelos de ecuaciones estructurales, fundamentados en el paquete AMOS. Aquí inició nuestro primer problema, dado que la Universidad no cuenta con el paquete AMOS para el desarrollo de estos modelos estadísticos y tuvimos que esperar hasta tener el dinero para comprar este material, para poder plantear el modelo y estimarlo.

Respecto a las dificultades encontradas en el desarrollo de la investigación, únicamente podemos mencionar una: la falta de paquetes informáticos de última generación para el análisis de datos. Este hecho es lamentable en una Universidad dedicada al desarrollo de la investigación y que no cuente con estas herramientas de apoyo. Las Universidades actuales dejaron hace muchos años las herramientas de análisis tradicionales porque se encontró que producían resultados erróneos, por tanto, no es lo mismo trabajar con los métodos tradicionales que con los emergentes. En este sentido, la única manera de resolver el problema fue utilizar los paquetes personales que este investigador ha adquirido. Sin embargo deja un sabor amargo saber que en una Universidad de prestigio, este tipo de herramientas no se encuentran para uso de sus profesionales.

4.5. Recomendaciones

A partir de los resultados obtenidos por esta investigación, se pueden derivar un conjunto de recomendaciones, que pueden ser utilizadas por los encargados de la administración e investigadores de nuestra universidad:

- La Universidad de Costa Rica propone un modelo teórico de docencia universitaria, que se asocia a los seis factores, los cuales han sido utilizados ampliamente por otras universidades a nivel internacional. Esto indica que se debe mantener la misma línea de trabajo por parte del C.E.A.
- Aunque los resultados empíricos no son concluyentes con la estructura teórica, se debe continuar trabajando sobre este tipo de modelos, dado que permiten determinar la validez de constructo, un componente que no se logra fácilmente.
- Es necesario la compra de paquetes informáticos de nueva generación, para el análisis de los datos. La Universidad no puede quedarse atrás de estas herramientas que permiten y mejoran las estimaciones.
- Se deberían tomar medidas asertivas para que los profesores universitarios conozcan los factores de docencia universitaria empleados por el C.E.A. y puedan prepararse adecuadamente para su futura evaluación.
- Como recomendación final, es sumamente importante considerar dentro del I.N.I.E. talleres de apoyo a los investigadores, profesores, estudiantes y público en general, acerca de las herramientas cuantitativas emergentes para el análisis de datos, debido a que existe un vacío muy grande sobre este tema. Además, es recomendable adquirir para el laboratorio del I.N.I.E. estos paquetes informáticos, a fin de que los investigadores puedan practicar y mejorar los trabajos de investigación cuantitativa.

4.6. Publicaciones

Se está trabajando en ello, en el anexo final se encuentra un artículo concluido para su publicación.

5. IMPACTO

Dentro del impacto de este proyecto se puede considerar el elemento científico, dado que nunca se había realizado un análisis de modelos de ecuaciones estructurales a ninguno de los instrumentos de evaluación del Centro de Evaluación Académica. Este hecho es un elemento particular de grandes dimensiones, pues comprueba la necesidad de nuestra institución de contar con instrumentos de evaluación que puedan cumplir con todas las normas establecidas respecto a la validez. En este sentido, no se puede confiar en ningún dato que no sea estudiado de manera profunda, dado que se ha comprobado que los análisis tradicionales aportan resultados con errores, por esa razón es necesario e importante continuar estudiando nuestros instrumentos de medida con los modelos de ecuaciones estructurales o métodos de última generación.

Otro impacto de la presente investigación, es el ámbito internacional debido a que con nuestro trabajo, pudimos apoyar al Banco Mundial, al Ministerio de Educación de Nicaragua y a la Universidad de Ingeniería de Nicaragua, porque ellos están trabajando sobre la construcción de instrumentos de evaluación docente y solicitaron nuestro apoyo para la estructura teórica. Además, también nos pidieron colaboración para el estudio de técnicas y métodos nuevos para el análisis de datos, en este sentido impartimos charla sobre los modelos de ecuaciones estructurales y los modelos jerárquicos lineales, ambos son técnicas emergentes para el análisis de datos.

6. BENEFICIO RECIBIDO PARA LA UNIVERSIDAD DE COSTA RICA

Dentro de los beneficios que recibe la institución, se consideran:

1.- Confianza de los profesores y administradores universitarios en la calidad de la prueba de evaluación docente y de los datos que aporte, pues alcanza buenos umbrales estadísticos.

2.- Reconocimiento por parte de los profesores, entidades nacionales e internacionales sobre la calidad y rigurosidad del instrumento aplicado. Por tanto, sería un reconocimiento de la gestión universitaria y de su preocupación por mejorar los procesos de evaluación y sus instrumentos, demostrando que están bien contruidos.

3.- Apoyo a otras universidades con deseos de medir la docencia universitaria con altos estándares de fiabilidad y validez, como en el caso del Banco Mundial, El Ministerio

de Educación de Nicaragua y la Universidad de Ingeniería de Nicaragua, organizaciones con las cuales estuvimos apoyando sobre el tema.

7. BIBLIOGRAFÍA CITADA Y CONSULTADA

Abalde, E.; De Salvador, X; González Carbanach, R. Y Muñoz C.; J.M. (1995). Análisis de la Evaluación de la Docencia Universitaria por los (as) alumnos (as) en la Universidad de la Coruña (1993-1994). **En Estudios de Investigación Educativa en Intervención Pedagógica**, pp. (289-292) Valencia: AIDIPE.

Abrami, P. C., Dickens, W. J., Perry, R. P., & Leventhal, L. (1980). Do teacher standards for assigning grades affect student evaluations of instruction. **Journal of Educational Psychology**, **72**, 107-118.

Abrami, P.C. (1989a). How Should We Use Student Ratings to Evaluate Teaching?. **Research in Higher Education**, **30** (2), 221-227.

Abrami, P.C.; Cohen, P.A. & D'apollonia, S. (1988). Implementation Problems in Meta- Analysis. **Review of Educational Research**, **58**, pp. 151-179.

Abrami, P.C.; D'apollonia, S. & Cohen, P. (1990). .Validity of Student Ratings of Instruction. What We Know and What We Do Not. **Journal of Educational Psychology**, **82** (2), pp.219-231.

Abrami.C. y D'apollonia, S. (1990b). .The dimensionality of ratings and their use in personnel decisions. M. Theall Y J. Franklins (Eds.), **In Student Ratings of Instruction. Issues for Improving Practice** (pp. 97-111). New Directions for Teaching and Learning.

Abrami, P.S.; D'apollonia, S. & Rosenfield. (1997). .The Dimensionality Of Student Ratings of Instruction. What We Know and What We do Not.. R.P. PRRY & J.C. SMART (Eds.) **In Effective Teaching in Higher Education. Research and Practice**, pp.321-367. New York. Agathon Press.

Acevedo Alvarez, R. & Rodríguez, N. M. (2006). Factores de sesgo asociados a la validez de la evaluación docente universitaria: un modelo jerárquico lineal. **Archivos Analíticos de Políticas Educativas**, **14** (34). Recuperado [27 marzo 2007] de <http://epaa.asu.edu/epaa/v14n34/>

Acevedo, R. y Fernández, M. J. (2004): “La percepción de los estudiantes universitarios en la medida de la competencia docente: validación de una escala”. En

Revista de Educación. N° 28 (2): 145-166. San José: Editorial de la Universidad de Costa Rica.

Albanese, M.A. (1991). .The validity of lecturer ratings by student and trained observers. **Academic Medicine**, **66** (1), pp. 26-28.

Aleamori, L. (1978). .Development and factorial validation of the Arizona Course/ Instructor Evaluation Questionnaire. **Educational and Psychological Measurement**, **38**, 1063-1067.

Aleamori, L. (1981). .Student ratings of instruction.. In J. MILLAMN (ed.), **Handbook of Teacher Evaluation**, pp.110-145. Newbury Park, CA. Sage.

Aleamori, L.M. (1999). .Student Rating myths versus research facts from 1924 to 1998.. **Journal of Personnel Evaluation in Education**, **13** (2), pp. 153-166.

Aleamori, L.M. y Hexner, P.Z. (1980). .A review of the research on student evaluation and a report on the effect of different sets of instructions on student course and instructor evaluation. **Instructional Science**, **9**, pp. 67-84.

Aleamori, L.M. y Yimer, M. (1974). Graduating Senior Ratings' Relationship to Colleague Rating Student Rating. **Research Productivity and Academic Rank in Rating Instructional Effectiveness (Research Report N°352)**. Urbana. University of Illinois, Office of Instructional Resources, Measurement and Research Division.

Anderson, K.H. y Siegfried, J.J. (1997). Gender differences in rating the teaching economics. **Eastern Economic Journal**, **23** (3), pp. 347-357.

Arubayi, E. (1987). .Improvement of Instruction and Teacher Effectiveness. Are Student Ratings Reliable and Valid?. **Higher Education**, **16**, pp. 267-288.

Barnes, L.B. y Barnes, M.W. (1993). .Academic discipline and generalizability of student evaluations of instruction. **Research in Higher Education**, **34** (2), pp. 135-149.

Bendig, A.W. (1953). .Relation of level of course achievement of student, instructor and course ratings in introductory psychology. **Educational and Psychological Measurement**, **13**, pp. 437-488.

Braskamp, L.A. & Ory, J.C. (1994). *Assessing Faculty Work. Enhancing Individual and Institutional Performance*. San Francisco. Jossey-Bass.

Braskamp, L.A.; Ory, J.C. & Pieper, D.M. (1981). Student Written Comments. Dimensions of Instructional Quality. **Journal of Educational Psychology**, **73** (1), pp. 75-70.

Broder, J.M. & Dorfman, J.H. (1994). Determinants of Teaching Quality; What's Important to Students?. **Research in Higher Education**, **35** (2), pp. 235-249.

Burdal, C.A. & Bardo, J.W. (1986). Measuring Student's Perceptions of Teaching. Dimensions of Evaluation. **Educational And Psychological Measurement**, **56**, pp. 63-79.

Campbell, D.T. y Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait – multimethod matrix. **Psychological Bulletin**, **56**, pp. 81-105.

Carson, B.H. (1999). Bad News in the Service of Good Teaching. Students Remember Ineffective Professors. **Journal On Excellence In College Teaching**, **19** (1), pp.91-105.

Casey, R.J.; Gentile, P. y Bigger, S.W. (1997). Teaching appraisal in higher education. an Australian perspective. **Higher Education**, **34** (3), 459-482.

Cashin, W. E. (1995). **Student Ratings Of Teaching. The Research Revisited**. IDEA Paper No. 32. Manhattan, KS: Kansas State University, Center for Faculty Evaluation and Development.

Cashin, W.E. (1988). **Student Ratings of teaching. A summary of the research**. IDEA. Paper No 20. Manhattan, KS. Kansas State University, Center Faculty Evaluation and Development.

CHACKO, T.I. (1983): "Student ratings of instruction: a function of grading standards". En **Educational Research Quarterly**, **83** (1), pp. 19-25.

Centra, J. (1977). Student Ratings of Instruction and Their Relationship to Student Learning. **American Educational Research Journal**, **14**, pp. 17-24.

Centra, J.A. (1972). **The Utility of Student Ratings for Instructional Improvement**. Princton, NJ. Educational Testing Services.

Centra, J.A. (1979). **Determining Faculty Effectiveness**. San Franciso: Jossey – Bass.

Centra, J.A. (1993). **Reflective Faculty Evaluation. nancing Teaching and Determining Faculty Effectiveness**. San Francisco: Jossey-Bass.

Cheng, Y. y Hoshower, L.B. (1998). Assessing student motivations to participate Teaching evaluations. an application of expectancy theory. **Issues in Accounting Education, 13** (3), pp. 531-549.

Coffey, M. & Gibbs, G. (2001). The Evaluation of the Student Evaluation of Educational Quality Questionnarie (SEEG) in U.K. Higher Education. **Assessment & Evaluation in Higher Education, Vol 26**, (1), pp. 89-93.

Cohen, P. A. (1981). Student ratings of instruction and student achievement. A meta-analysis of multisection validity studies. **Review of Educational Research, 51**, 281-309.

Cohen, P.A. (1983). A selective review of the validity of student ratings of teaching. **Journal of Higher Education, 54**, 448-458.

Cohen, P.A. (1989). Do grades influence students' evaluation of clinical courses?. **Journal of Dental Education, 53** (4), pp. 238-240.

Costin, F. (1968). **Survey of Opinions About Lectures**. University of Illinois: Department of Psychology.

Costin, F.; Greenough, W.T. y Menges, R.J. (1971). Student Ratings Of College Teaching. Reliability, Validity and Usefulness. **Review Of Educational Research, 41**, 511-535.

Craton, P. y Smith, R. A. (1990). Reconsidering the Unit of Analysis. A Model of Student Ratings of Instruction. **Journal of Educational Psychology, 82** (2) pp. 207-212.

Cruse, D. B. (1987). Student Evaluation and the University Professor. **Higher Education, 15** (6), pp. 723-737.

D'apollonia, S. & Abrami, P. C. (1997). Navigating Student Ratings of Instruction. **American Psychologist**, **51** (11), pp. 1198-1208

De Neve, H. M. F. y Janssen, P. J. (1982). Validity of Student Evaluation of Instruction. **Higher Education**, **11** (5), pp. 543-552.

Dickinson, D. J. (1990). The relationship between ratings of teacher performance and student learning. **Contemporary Educational Psychology**, **15**, pp. 142-151.

Dowell, D. A. , & Neal, J. A. (1982). A selective view of the validity of student ratings of teaching. **Journal of Higher Education**, **53**, 51-62.

Doyle, K. O. (1975). **Student Evaluation of Instruction**. Lexington, MA: Lexington Books.

Drews, D. R. ; Burroughs, W. J. Y Nokovich, D. A. (1987). Teacher self ratings as a validity criterion for student evaluation. **Teaching of Psychology**, **14** (1), pp. 129-143.

Drucker, A. J. Y Remmers, H. H. (1950). Do Alumni and Students Differ in Their Attitudes Toward Instructors?. **Purdue University Studies in Higher Education**, **70**, pp. 62-64.

Drucker, A. J. Y Remmers, H. H. (1951). Do Alumni and Students Differ in Their Attitudes Toward Instructors?. **Journal of Educational Psychology**, **42**, pp. 129-143.

Erdle, S.; Murray, H.G. & Rushton, J.P (1985): "Personality, Classroom Behavior and Student Ratings of College Teaching Effectiveness: A Path Analysis". En **Journal of Educational Psychology**, **77** (4), pp. 394-407.

Feldman, K. A. (1977). Consistency and Variability among College Students in Rating Their Teachers and Courses. **Research in Higher Education**, **6** (2), pp. 223-274.

Feldman, K. A. (1978). Course characteristics and college students' ratings of their teachers. What are know and what we don't know. **Research in Higher Education**, **9** (2), pp. 199-242.

Feldman, K. A. (1979). The Significance of Circumstances for College Students ratings of Their Teachers and Courses. **Research in Higher Education**, **10** (2), pp. 149-172.

Feldman, K. A. (1984). Class size and college students' evaluation of teachers and courses. a closer look. **Research in Higher Education**, **21** (11), pp. 45-116.

Feldman, K. A. (1989a). Instructional effectiveness of college teachers as judged by teachers themselves, current and former student, colleagues, administrators and external (neutral) observers. **Research in Higher Education**, **30** (2), pp. 137-194.

Feldman, K. A. (1989b). The association between student ratings of specific instructional dimensions and student achievement. refining and extending the synthesis of data from multisection validity studies. **Research in Higher Education**, **30**, pp. 583-645.

Feldman, K. A. (1997). Identifying Exemplary Teachers and Teaching. Evidence from Student Ratings. R. P. En PERRY & J. C. SMART (eds.), **Effective teaching in Higher Education. Research and Practice**, . (pp. 368-395). Bronx, N. Y: Agathon.

Fernández, J. ; Mateo, M. & Muñiz, J. (1998). Is There Relationship Between Class Size and Student Ratings of Teacher Quality?. **Educational And Psychological Measurement**, **58** (August), 596-604

Franklin, J. & Theall, M. (1989). **Who read ratings. Knowledge, attitude, and practice of users of student ratings of instruction**. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Franklin, J. & Theall, M. (1989). **Who read ratings. Knowledge, attitude, and practice of users of student ratings of instruction**. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Frey, P. W. (1973). Student Ratings of Teaching. Validity of Several Rating Factors. **Science**, **182**, pp. 83-85

García Ramos, J. M. (1999^a). Análisis multirasgo- multimétodo en la validación de instrumentos de medida para la evaluación de la calidad docente en instituciones universitarias. **Revista Española de Pedagogía**, **214**, pp. 417-444.

García Ramos, J. M. (1997). Valoración De La Competencia Docente Del Profesor Universitario. Una Aproximación Empírica. **Revista Complutense De Educación**, **8** (2), pp. 81-108

García Ramos, J. M. y Congosto Luna, E. (1996). **Un Modelo de Evaluación Institucional en la Universidad**. Salamanca: Studia Pedagógica

Gigliotti, R. Y Buchtel, F. (1990). Attributional Bias and Course Evaluations . **Journal of Educational Psychology**, **82** (2), pp. 341-351.

Gillmore, G. (1973). **Estimates of Reliability Coefficients for Items and Subscales of the Illinois Courses Evaluation Questionnaire**. (Research Report N°341). Urbana: University of Illinois, Office of Instructional Resources, Measurement, and Research Division.

Gillmore, G. M. ; Kane, M. T. & Maccarato, R. W. (1978). The Generalizability of Student Ratings of Instruction. Estimation of the Teacher And Course Components. **Journal of Educational Measurement**, **15**, pp. 1-13.

Gilmore, G. (1984). Student ratings as a factor in faculty employment decisions and periodic review. **Journal of College and University Law**, **10**, 557-576 .

Goddard, R; Hoy, W & Woolfolk, A. (2001). Collective Teacher Efficacy. Its Meaning, Measure, and Impact on Student Archievement. **Collective Teacher Efficacy**. Manuscrito sometido para publicación. 1-40.

Goldman, L. (1993). On erosion of education and the eroding foundation of teacher education (or why we should nor take student evaluation of faculty seriously). **Teacher Quarterly**, **20** (2), pp. 57-64.

Greenwald, A. G. (1997b). Validity Concern and Usefulness of Student Ratings Of Instruction. **American Psychologist** , **51** (11), 1182-1186.

Greenwald, A. G. Y Gillmore, G. M. (1997). Grading Leniency is a Removable Contaminant of Student Ratings. **American Psychologist** , **51** (11), 1209-1217.

Guthrie, E. R. (1954). **The Evaluation of Teaching**. A Progress Report. Seattle: University of Washington.

Harrison, P. D. ; Ryan, J. M. Y Moore, P. S. (1996). College student's self-insight and common implicit theories in the ratings of teaching effectiveness. **Journal of Educational Psychology**, **88** (4), pp. 775-782.

Hativa, N. (1996). University instructors' rating profiles. Stability over time, and disciplinary differences. **Research in Higher Education**, **37** (3), pp. 341-365.

Hativa, N. y Raviv, A. (1993). Using a single score for summative teacher evaluation by students. **Research in Higher Education**, **34** (5), pp. 625-646.

Hepworth, D. Y Oviatt, B. E. (1985). Using student course evaluations. findings, issues and recommendations. **Journal of Social Work Education**, **21** (3), pp. 105-112.

Hilton, H. (1993). Realibility and Validity of Student Evaluation. Testing Models versus Survey Research Model, p. s. **Political Science & Politics**, **26**, pp. 562-569.

Hogan, T. P. (1973). Similarity of student ratings across instructors, courses and time. **Research in Higher Education**, **1**, pp. 149-154.

Holmes, D. S. (1972). Effects of grades and disconfirmed grade expectancies on students' evaluations of their instructor. **Journal of Educational Psychology**, **63**, 130-133.

Howard, G. S. , Conway, C. G, & Maxwell, S. E. (1985). Construct validity of measures of college teaching effectiveness. **Journal of Educational Psychology**, **77**, 187-196.

Howard, G. S. Y Maxwell, S. E. (1980). Correlation between student satisfaction and grades. a case of mistaken causation?. **Journal of Educational Philosophy**, **72** (4), pp. 810-820.

Howell, A. J. Y Symbaluk, D. (2001). Published Student Ratings of Instruction. Revealing and Reconciling the Views of Student and Faculty. **Journal of Educational Psychology**, **(4)**, 790-796

Jornet, J. M. ; González Such, J. Y Pérez Carbonell, A. (1995) . Evaluación De la Actividad Universitaria. G. Quintás (eds.), En **Reforma y Evaluación de la Universidad**. Valencia: Servei de Publicacions de la Universitat de Valencia. pp. 189-244.

Kikut, Lorena (2003). **Exploración de la confiabilidad y validez del cuestionario de evaluación docente**. San José : Universidad de Costa Rica.

Koblitz, . N. (1990). Are student ratings unfair to women?. **Newsletter of the Association for Women in Mathematics, 20**, pp. 17-19.

Koermer, C. D. Y Petelle, J. L. (1991). Expectancy violation and student rating of instruction. **Communication Quarterly, 39** (4), pp. 341-350.

Kolitch, E. & Dean, A. V. (1999). Student Ratings of Instruction en the U. S. A. Hidden Assumptions and Missing Conceptions About . Good. Teaching. **Studies in Higher Education, Vol. 24**, (1), pp. 27-42.

Koon, J. Y Murray, H. G. (1995). Using multiple outcomes to validate student ratings of overall teacher effectiveness. **Journal of Higher Education. 66** (1), pp. 61-81.

Lin, W. Y. ; Watkins, D. Y Meng, Q. M. (1995). Student's Evaluation of University teaching. A China perspective. **Higher Education and Research y Development, 14** (1), pp. 61-74.

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

Mahmound, M. M. (1991). Descriptive models of student decision behaviour in evaluation of higher education. **Assessment and Evaluation in Higher Education, 16** (2), pp. 133-148.

Marques, T. E. ; Lane, D. M. Y Dorfman, P. W. (1979). Toward the development of a sistem for instructional evaluation. Is there consensus regarding what constitutes effective teaching?. **Journal of Educational Psychology, 71**, pp. 840-849.

Marsh, H. & Roche, L. R. (1997). Making Students' Evaluation of Teaching Effectiveness Effective. The Critical Issues of Validity, Bias, and Utility. **American Psychologist, Vol. 52**, 11, 1187-1197.

Marsh, H. (1982). SEEQ. a Realiable, Valid, and Useful Instrument for Collecting Students Evaluation of University Teaching. **British Journal of Psychology, 52**, pp. 77-95.

Marsh H. (1984). Student's evaluation of university teaching; dimensionality, reliability, validity, potential biases and utility. **Journal of Educational Psychology**, 76 (5), 707-754.

Marsh, H. (1987a). Students' evaluation of university teaching; Research findings, methodological issues, and directions for future research. **International Journal of Educational Research**, 11, pp. 253-288.

Marsh, H. (1987b). Student Evaluations of Teaching. M. J. DUNKINS (eds.), **The International Encyclopedia of Teaching and Teacher Evaluation**. (pp. 181-187) Oxford: Pergamon Press

Marsh, H. y Bailey, M. (1993). Multidimensional Students' Evaluations of Teaching Effectiveness. **Journal of Higher Education**, 64 (1), pp. 1-18.

Marsh, H. A. y Overall, J. U. (1979b). Validity of student's evaluation of teaching. A comparison with self evaluations by teaching assistants, undergraduate faculty, and graduate faculty. Paper presented Annual Meeting of the American Educational Research Association, San Francisco.

Marsh, H. W. (1977). The validity of students' evaluations of instructors independently nominated as best and worst teacher by graduating senior. **American Educational Research Journal**, 14, pp. 441-447.

Marsh, H. W. (1980). The influence of student, course, and instructor characteristics on evaluations of university teaching. **American Educational Research Journal**, 17, 219-237.

Marsh, H. W. (1992b). **A Longitudinal Perspective of Student's Evaluations of University Teaching. Ratings of The Same Teacher over a 13 Year Period**. Documento presentado en Annual Meeting of the American Educational Research Association (p. 18) San Francisco, Ca. Abril.

Marsh, H. W. (2001). Distinguishing between good (useful) and bad workloads on Students' Evaluations of teaching. **American Educational Research**, 38 (1), pp. 183-212.

Marsh, H. W. y Dunkin, M. J. (1992). Students' Evaluation of University Teaching. A Multidimensional Perspective. J. SMART (ed.) **Higher Education. Handbook of Theory and Research**. (pp. 143-223) New York: Agathon.

Marsh, H. W. y Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept. First and higher order factor models and their invariance across groups. **Psychological Bulletin**, **97**, pp. 562-582.

Marsh, H. W. y Overall, J. U. (1981). The relative influence of course level, course type, and instructor on students' evaluations of college teaching. **American Educational Research**, **18**, pp. 103-112.

Marsh, H. W. y Roche, L. (1992). **The Use of Student's Evaluations of University Teaching To Improve Teaching Effectiveness**. Canberra: Australian Government Publishing Services.

Marsh, H. W. y Roche, L. A. (1994). The Use of Student Evaluations of University Teaching in Different Settings. The Applicability Paradigm. **Australian Journal of Education**, **36** (3), pp. 278-300.

Marsh, H. W. y Roche, L. A. (2000). Effects of grading leniency and low workload on students' evaluation of teaching. popular myth, bias, validity or innocent bystanders?. **Journal of Educational Psychology**, **92** (1), pp. 202-228.

Marsh, H. W. ; Balla, J. R. y Mcdonald, R. P. (1988). Goodness of Fit Indices in Confirmatory Factor Analysis. The Effects of Sample Size. **Psychological Bulletin**, **103**, pp. 391-410.

Marsh, H. W. ; Hau, K. T. ; Chung, C. M. & Siu, T. L. P. (1997). Students' s Evaluations of University Teaching. Chinese Version of The Student's Evaluations of Educational Quality Instrument. **Journal of Educational Psychology**, **89** (3), pp. 568-572.

Marsh, H. & Roche, L. R. (1997b). Making Students's Evaluation of Teaching Effectiveness Effective. The Critical Issues of Validity, Bias, and Utility. **American Psychologist**, **Vol. 52** (11), pp. 1187-1197.

Marsh, H. W. ; Touron, J. y Wheeler, B. (1985). Students' s Evaluations of University Instructor. The Applicability of American Instrument in a Spanish Setting. **Teaching and Teacher Education**, **1**, pp. 123-138.

Mccallum, L. W. (1984). A meta-analysis of course evaluation data and its use in the tenure decision. **Research in Higher Education**, **21**, 150-158.

McKeachie, W. J. (1979). Student ratings of faculty. A reprise. **Academe**, **65**, pp. 384-397.

McKeachie, W. J. (1987). Instructional evaluation. current issues and possible improvement. **Journal of Higher Education**, **58** (3), pp. 344-350.

McKeachie, W. J. (1990). Research on College Teaching. The Historical Background. **Journal of Educational Psychology**, **82** (2), pp. 189-200.

McKeachie, W. J. (1997). Student Ratings. The Validity of Use. **American Psychologist**, **52** (11), pp. 1218-1225.

McKeachie, W. J. ; LIN, Y. G. y MENDELSON, C. N. (1978). A small study assessing teacher effectiveness. Does learning last?. **Contemporary Educational Psychology**, **3**, pp. 352-357.

Meeth, L. R. (1976). The stateless art of teaching evaluation. report on teaching. Vol. 2, Change, 8, pp. 3-5.

Menges, R. J. (1991). The real world of teaching improvement. A faculty perspective. M THEALL & FRANKLIN (Eds.) **Effective Practices for Improving Teaching, New Directions for Teaching and Learning, Vol. 48**, (pp. 21-37). San Francisco: Jossey-Bass.

Miller, R. I. (1987). **Evaluation Faculty for Promotion and Tenure**. San Francisco: Jossey-Bass.

Miller, S. (1984). Student rating scales for tenure and promotion. **Improving College and University Teaching**, **32** (2), pp. 87-90

Millman, J. (1981). **Handbook of Teacher Evaluation**. Beverly Hills, CA: Sage.

Monroe, C. y Borzi, M. G. (1989). Methodological issues regarding student evaluation of teacher. A pilot study. **ACA Bulletin**, **70**, pp. 73-79.

Moses, I. (1986). Self and Student evaluation of academic staff. **Assessment and Evaluation in Higher Education**, pp. 76-78.

Muñiz, J. ; García, A. y Virgos, J. M. (1991). Escala de la Universidad de Oviedo para la evaluación del profesorado. **Psicothema**, **3** (2), pp. 269-281.

Murray, H. G.; Rushton, P. & Paunonen, S. V. (1990). Teacher Personality Traits and Student Instructional ratings in Six Types of University Courses. **Journal of Educational Psychology**, **82** (2), pp. 250-261.

Nimmer, J. G. y Stone, E. F. (1991). Effects of grading practices and time of rating on student ratings of faculty performance and student learning. **Research in Higher Education**, **32** (2), pp. 195-215.

Noser, T. C. ; Manakyan, H. y Tanner, J. R. (1996). Research productivity and perceived teaching effectiveness. A survey of economic faculty. **Research in Higher Education**, **37** (3), pp. 299-321.

O'connell, D. Q. y Dickenson, D. J. (1993). Student ratings of instruction as a function of testing conditions and perceptions amount learned. **Journal of Research and Development in Education**, **27** (1), pp. 18-23.

Overall, J. U. & Marsh, H. W. (1980). Students' s Evaluations of Instruction. A Longitudinal Study of Their Stability. **Journal of Educational Psychology**, **72**, pp. 321-325.

Palchik, N. S. (1988). Student assessment of teaching effectiveness in a multi-instructor course for multidisciplinary health professional student. **Evaluation and the Health Professions**, **11** (1), pp. 55-73.

Powell, R. W. (1977). Grades, learning, and student evaluation of instruction. **Research in Higher Education**, **7**, 193-205.

Prave, R. S. y Bavril, G. L. (1993). Instructor rating. controlling for bias from initial student interest. **Journal of Educational for Business**, **68** (2), pp. 362-366.

Prosser, M. & Trigwell, K. (1990). How will Future Academic be Evaluated? Using Student Study Strategies to **Check the Validity of Student Evaluations of Teaching Courses**. G. MULLIS (Ed)

Ramsden, P. (1991). A performance indicator of teaching quality in higher education. the course experience Questionnaire. **Studies in Higher Education, 16**, pp. 129-150.

Rindermann, H. & Schofield, N. (2001). Generalizability of Multidimensional Student Rating of University Instruction Across Courses and Teacher. **Research in Higher Education. Vol. 42**, 4, pp. 377-400.

Rodin, M. & Rodin, B (1972). Student Evaluation of Teachers. **Science, 177**, pp. 1164-1166.

Rutland, P. (1990). Some considerations regarding teaching evaluations. **Political Science Teacher, 3**, pp. 1-2.

Ryan, J. M. & Harrison, P. D. (1995). The Relationship Between Individual Instructional Characteristic and the Overall Assessment of Teaching Effectiveness Across Different Instructional Context. **Research In Higher Education, Vol. 36**, No 5, pp. 577-594

Ryans, D. G. (1960) **Characteristics of Teachers**. Washington, D. C.: American Council on Education.

Salvador, L. (1990). Los Docentes Universitarios Exitosos Desde La Perspectiva del Alumno. Su Caracterización Psicopedagógica. Tesis Doctoral. Universidad de Salamanca. España.

Saroyan, A. & Amundsen, Ch. (2001). Evaluating University teaching. Time to Take Stock. **Assessment & Evaluation in Higher Education, Vol 26**, (4), pp. 341-353

Seldin, P. (1984). Changing Practices in Faculty Evaluation. A Critical assessment and Recommendations for Improvement. San Francisco: Josser – Bass.

Seldin, P. (1993a). The use ad abuse of students ratings of professors. **The Chronicle of Higher Education**, p. 40

Seldin, P. (1993b). **Successful Use Of Teaching Portafolios**. Bolton, MA: Anker Publishing Co.

Shadish, W. (1998). **Some Evaluation Questions. Practical Assessment, Research & Evaluation, 6 (3)**. Recuperado el 14 de enero de 2001 de <http://ericae.net/pare/getvn.asp?v=6&n=3>].

Snyder, C. R. , & Clair, M. (1976). Effects of expected and obtained grades on teacher evaluation and attribution of performance. **Journal of Educational Psychology, 68**, 75-82.

Spencer, P. A. Y Flyr, M. L. (1992). **The Formal Evaluation As Ac Impetus To Classroom Change. Myths or Reality?**. Research/ Technical Report, Riverside, CA.

Tagomori, H. & Bishop, L. (1995). Student Evaluation of Teaching. Flaw Instruments. Thought and Action. **The National Education Association Higher Education Journal, 11**, pp. 63-78.

Tatro, C. N. (1995). Gender effect on student evaluations of faculty. **Journal of Research and Development in Education, 28 (3)**, pp. 169-173.

Tejedor, F. y Montero, L. (1990). Indicadores de la Calidad Docente para La Evaluación del Profesor Universitario. **Revista Española De Pedagogía, año XLVIII, N° 186**, mayo-agosto, pp. 259-279.

Timpson, W. W. Y Andrew, D. (1997). Rethinking student evaluation and the improvement of teaching. instrument for change at the University of Queensland. **Studies in Higher Education, 22 (1)**, pp. 55-65.

Ting, K. F. (2001). A Multilevel Perspective On Student Ratings of Instruction. Lessons From the Chinese Experience. **Research in Higher Education. Vol. 41, 5**, pp. 637-653.

Trigwell, K. & Prosser, M. (1996). Changing Approaches to Teaching. A Relational Perspective. **Studies in Higher Education, 21**, pp. 275-284.

Vandewalle (1997). Development and validation of a work domain goal orientation instrument. **Educational and Psychological Measurement, 57 (6)**, pp. 995-1015.

Vasta, R. , & Sarmiento, R. F. (1979). Liberal grading improves evaluations but not performance. **Journal of Educational Psychology, 71**, 207-211.

Villa, A. Y Morales, P. (1993). **La Evaluación Del Profesor. Una Visión De Los Principales Problemas Y Enfoques De Diversos Contextos.** Vitoria: Departamento de Educación, Universidades e Investigación. Gobierno Vasco.

Vu, T. R. ; Marrito, D. J. ; Stratos, G. A. Y Litzelman, D. K. (1997). Prioritizing areas for faculty development of clinical teachers by using student evaluations for evidence- based decisions. **Academic Medicine**,**72** (10), pp. 57-59.

Wachtel, H. K. (1998). Student Evaluation of College Teaching Effectiveness. A Brief Review. **Assessment & Evaluation In Higher Education**, **23** (June), 191-211.

Weinbach, R. W. (1988). Manipulation of student evaluations. No laughing Mater. *Journal of Social Work Education*, 24 (1), pp. 37-34.

Wigington, H. ; Tollefson, N. Y Rodríguez, E. (1989). Student's ratings of instructor visited. Interactions among class and instructor variables. **Research in Higher Education**, **30** (3), pp. 331-334.

Wilkerson, D. ; Rogers, M. A. Y Maughan, R. (2000). Validation of Student, Principal, and Self Ratings in 360 Feedback for Teacher Evaluation. **Journal of Personnel Evaluation in Education**, **14** (2), pp. 179-192

Wilson, K. L.; Lizzio, A. & Ramsden, P. (1997). The Development , Validation, and Application of the Course Experience Questionnaire. **Studies in Higher Education**, **22** (1) , pp. 33-52.

Worthington, A. G. , & Wong, P. T. P. (1979). Effects of earned and assigned grades on student evaluations of an instructor. **Journal of Educational Psychology**, **71**, 764-775.

Worthington, A. C. (2002). The impact of student perception and characteristics on teaching evaluation. a case study in finance education. **Assessment & Evaluation in Higher Education**, **27** (1), pp. 49-64.

Zoller, U. (1992). Faculty Teaching Performance Evaluation in Higher Science Education. Issues and Implications. **Science Education**, **76** (6), pp. 673-684.

ANEXOS

ANEXO N° 1 Síntesis: La Evaluación Docente

VISIÓN PANORÁMICA

Durante los 80 y 90, ocurrieron *rápidos cambios* en la mayoría de aspectos de nuestra sociedad, tales cambios en el medio ambiente, sobre todo, afectaron la escuela, y los profesores tuvieron que encarar diferentes tipos de problemas, inciertos y desafiantes. Como es de esperar, las metas en educación parecen complejas: las tareas tienen más demanda; las expectativas del público son más diversas; la rendición de cuentas toma auge. Pero en el siglo XXI el ambiente de la educación podría ser más cambiante y demandante, porque el desarrollo de la tecnología, la economía y el ambiente político es rápido y drástico. Esto supone que los profesores, en la era de cambios, están requiriendo la obligación de ampliar sus roles y responsabilidades como desarrolladores del currículo, del nuevo profesor mentor, facilitador del desarrollo del grupo, la acción de la investigación, el equipo líder. En este sentido las nuevas responsabilidades envuelven tanto a los profesores, los padres, los estudiantes, así como a los organismos públicos, los encargados de la toma de decisiones y a los miembros de la administración universitaria.

EL CONCEPTO

Nosotros entendemos que la *evaluación de la competencia docente universitaria* es un proceso en constante desplazamiento entre la acción, reflexión, construcción y reconstrucción: no como una actividad puntual, que ha de realizarse en un momento determinado y que concluye, una vez recogida y analizada la información. Como señala Tejedor (1990) es una actividad orientada fundamentalmente a la estimación del *nivel de calidad* de la enseñanza universitaria, a fin de contribuir progresivamente a su *mejora*. En esta definición, encontramos dos elementos destacados constantemente por los investigadores: el primero, discutido por Feldman (1997), consiste en incrementar el énfasis

en la calidad de la enseñanza para conseguir mejores resultados y el segundo, destacan Saroyan y Amundsen (2001), es un recurso que puede contribuir con la mejora de los docentes universitarios, por ende su enseñanza.

De la Orden (1992), la define como un *proceso sistemático* de recolección, análisis e interpretación de información relevante de la labor docente y la formulación de un juicio de valor sobre su adecuación a un criterio o patrón, como base para la toma de decisiones acerca de tal faceta.

En este sentido, la evaluación de la competencia docente universitaria ha de *proporcionar un conocimiento* de la situación de partida y del grado de coherencia de sus elementos integrantes. Tras su valoración, ha de elaborarse un programa de mejora de sus aspectos susceptibles. Una tarea como ésta, exige una serie de etapas que van desde la planificación previa, definición de objetivos educativos, determinación de los sistemas de referencia, hasta las técnicas de registro, análisis y la misma redacción del informe.

LOS PROPÓSITOS

Villa y Morales (1993), en un buen análisis y clasificación de los propósitos de evaluación, afirman que éstos pueden *referirse a distintos ámbitos* educativos en los que se encuentra envuelta la acción del docente. En opinión de los autores, se distinguen seis propósitos, relacionados con el profesor, el estudiante, la dirección del centro, la sociedad, la institución y la investigación. A continuación, se describen aquellos que tienen relación con cada uno de los implicados:

El profesor

- Ayudar a mejorar su acción didáctica
- Proporcionarle feedback (puntos débiles y fuertes)
- Favorecer el perfeccionamiento docente en destrezas y habilidades instructivas
- Ayudar a ver puntos oscuros, desconocidos para él

Alumnos

- Orientar al alumno en la selección de áreas y disciplinas
- Proporcionar información a los alumnos para seleccionar cursos e instructores

Dirección de centros

- Proporcionar datos sobre los que poder tomar decisiones con propósitos personales
- Tomas de decisión sobre contratación, promoción y retribución del profesorado.

Sociedad

- Informar a la sociedad sobre el rendimiento del profesorado y de las instituciones educativas
- Rendir cuentas ante la sociedad

Institución

- Elaborar y mantener un inventario de recursos personales para reajustes, perfeccionamiento, etc.
- Evaluar y mejorar el currículo, la programación y la acción departamental

Investigación

- Proporcionar datos para la investigación sobre la enseñanza, programas, acción docente y aprendizaje.

Exposiciones detalladas sobre cada uno de estos aspectos se pueden ver en Natriello (1990), quién, además de describir las funciones pretendidas por los diferentes implicados, detalla también las consecuencias intencionadas y no intencionadas que se producen con la evaluación.

FINALIDADES DE LA EVALUACIÓN

La evaluación formativa se caracteriza por

- a) La información es recibida únicamente por el propio profesor.
- b) La información puede recogerse frecuentemente e informalmente
- c) La evaluación orientada al propio desarrollo personal es muy útil a largo plazo
- d) La información compartida con otro a menudo aumenta la utilidad de la información.

La sumativa

- Existe una paradoja sobre la necesidad de rendir cuentas a la sociedad y la libertad del profesorado para expresarse.
- Debe existir utilidad y relación entre la actuación docente y su evaluación y de alguna forma gratificación por la calidad docente.
- Tener en cuenta el mérito del profesor como su valor para la institución.
- Esta evaluación es una manera expresiva de comunicar las expectativas de la institución
- Tiene que tener credibilidad tanto para el docente como para los administradores
- La información utilizada debe tener suficiente calidad
- Debe basarse en principios y prácticas legales
- La institución puede examinar procedimientos alternativos de evaluación que puedan ser provechosos para ella.

LOS USUARIOS

El primer usuario es *el académico*, que requiere la evaluación de su curso y de su enseñanza. Generalmente tiene dos propósitos: el primero consiste en la necesidad de saber o tener un feedback de sus estudiantes, contar con un elemento que le permita saber cómo ven ellos la enseñanza que reciben; ésta es una visión que le sirve para incrementar y mejorar su enseñanza en aquellos elementos que considere necesarios en el futuro. El autor destaca que es importante anotar, que, dentro de las políticas de evaluación de la Universidad, solamente este usuario, el profesor, puede comenzar una evaluación de su enseñanza.

El segundo según Rindermann y Schofield (2001) es el uso que el académico brinde a los resultados, ya sea en su carrera, para comparar los resultados de evaluación con su propio rendimiento pasado o con el propio rendimiento de sus compañeros. En fin, para otras actividades que no sean necesariamente la retroalimentación.

Otro tipo de usuario son los académicos miembros del comité de selección, trabajo y promoción. Éstos necesitan de formas e instrumentos para juzgar la aplicación de la enseñanza de calidad. De hecho, ellos necesitan contar con criterios e indicadores que le permitan la comparación entre evaluados y su enseñanza de calidad, para tomar decisiones sobre la carrera del docente relacionada con aumento de salario, cambio de rango, continuidad laboral. Para Yorke (1995) esas formas de comparación sobre decisiones básicas afectan enormemente la carrera de los docentes, porque se relacionan directamente con su remuneración y algunas veces, con la financiación del departamento al que pertenecen, por eso, los juicios de los académicos miembros del comité de selección son muy importantes y necesitan criterios muy claros, que les permitan emitir veredictos certeros, para valorar con precisión la acción docente.

El *tercer grupo* de usuarios de evaluación docente son los promotores de personal, los cuales están llamados a asesorar a los miembros de la comunidad universitaria en todos los aspectos de la evaluación, incluyendo la importante área de interpretación de la encuesta, su implicación en las decisiones acerca de la enseñanza de calidad y mejora.

Finalmente, Franklin y Theall (1989) definen que el *cuarto grupo de usuarios* son los estudiantes. Generalmente a ellos les gusta conocer que el tiempo que invierten en completar las encuestas se relaciona con los cambios en el contexto de enseñanza y que su genuina visión está teniendo impacto en la calidad de la enseñanza que reciben. Aparte de esto, también emplean los resultados de la evaluación como referente a la selección de sus futuros cursos y profesores. Sin embargo, dada la confidencialidad de los resultados de evaluación en la universidad, los estudiantes generalmente no tienen ese acceso a la información sobre las encuestas de personal y distan mucho de saber los cursos ofrecidos por profesores con bajas calificaciones en el departamento

ACCIONES BÁSICAS PARA PROMOVER EL PROCESO DE EVALUACIÓN

Para alcanzar una evaluación con éxito de la competencia docente universitaria, el sistema evaluativo deberá construirse a partir de un equilibrio dinámico entre la mejora de la escuela y la mejora del profesorado; para ello será necesario tener en cuenta, según Stronge (1997) (citando a Glatthorn, 1984; Conley, 1987; Harris, 1987 y Mcgreat, 1988) las siguientes acciones prioritarias: establecer objetivos conjuntos, comunicación total, creación de un clima evaluador, garantía técnica de evaluación y uso de varias fuentes de datos:

- a) Establecimiento de *objetivos mutuamente beneficiosos*: éstos van a constituir las metas por conseguir, deben ser valorados y percibidos como importantes tanto por los profesores como por la institución. Además, el establecimiento de metas y objetivos deben ser aceptados por todos los involucrados, esto es un asunto de consenso y no de imposición, por tanto ha de considerarse la opinión de todos los involucrados y sus necesidades.

- b) Una *comunicación sistemática*: la transparencia de los procesos y la posibilidad de interacción, se fundamenta en políticas eficientes de información, por tanto, cada aspecto clave del proceso de evaluación deberá ir acompañado del correspondiente acto informativo, que permita fluir la información, llegando rápidamente a todos los miembros del claustro. Los canales de comunicación han de ser abiertos, permanentes y claros, esto permite fluidez a la información.

- c) Creación de un *clima favorecedor de la evaluación*: La evaluación sólo puede resultar absolutamente eficiente si se lleva a cabo desde un clima adecuado donde la confianza entre las partes implicadas, la honestidad y la transparencia de actuaciones, sean la norma y no la excepción. La consecución de estos climas posibilita, en la medida en que se siguen los pasos establecidos, participación en el establecimiento de los objetivos, comunicación continua y fluida en las diferentes fases del proceso, projesión en la aplicación de las técnicas, honestidad en los juicios y enfoques pro-activos en los procesos de mejora.

Garantía de aplicación *técnica del sistema evaluativo*: Cada uno de los aspectos que conforman la aplicación técnica de los procesos evaluativos, se realiza de acuerdo a las condiciones técnicas de calidad y precisión exigidas

Ámbito de la evaluación de la competencia docente

Prácticamente, uno de los primeros pasos en la evaluación de la competencia docente universitaria, radica en detenerse a considerar *qué* se desea evaluar. Exposiciones detalladas sobre este tema pueden consultarse en la generosa gama de investigaciones realizadas hasta el momento, siendo especialmente recomendable la de Medley, Coker y Soar (1984), quienes consideran la conveniencia de delimitar una serie de conceptos que a

menudo se intercambian y se utilizan como sinónimos. Estos autores distinguen cuatro términos:

- a) Aptitudes docentes (*Teacher Competency*): conocimientos, habilidades o valores específicos, que el profesor puede o no puede poseer y que se consideran importantes para ser un buen profesor.
- b) Competencia docente (*Teacher Competence*): se considera el repertorio de competencias que posee el profesor. Cuanto mayor sea este repertorio, más competente se dice que es el profesor.
- c) Actuación docente (*Teacher Performance*): Lo que hace el profesor en su trabajo, definido en términos de la conducta competente del profesor bajo condiciones específicas, dependerá, en parte, de lo competente que sea, y en parte, de la situación y contexto en el que se realiza la acción, así como también de la capacidad para aplicar sus competencias en u momento dado.
- d) Efectividad docente (*Teacher Effectiveness*): los resultados que un profesor logra, definidos en términos de lo que hace el alumno o su rendimiento.

Tal como indican Medley, Coker y Soar (1984), existen cinco momentos o situaciones importantes en la carrera profesional de un profesor en la que éste pueda ser evaluado con distintos propósitos y en términos de diferentes dimensiones de enseñanza. Desde esta propuesta, la evaluación del profesor debería englobar distintas dimensiones en las diversas situaciones por las que transcurre su vida profesional.

ANEXO N° 2 Factores De Evaluación Docente

Factores identificados por los estudiantes asociados con la competencia docente

La búsqueda por los factores asociados con la competencia docente, desde la visión de los alumnos, ha sido la fuente de trabajo de numerosos estudios e investigaciones, apoyados en diferentes posturas teóricas las cuales les han permitido identificar los predictores de evaluación de la acción universitaria.

En esta incesante búsqueda se pueden observar trabajos muy importantes como los de Feldman (1976^a, 1976b, 1987, 1989) que destaca la existencia de 14 a 20 dimensiones analizadas por los investigadores. En su estudio el autor esquematiza estas investigaciones e identifica varios constructos¹, agrupados concisamente en tres factores de primer orden y relacionados con el rol del profesor: 1) el presentador (un hombre actor o presentador, presentador de materiales), 2) facilitador del aprendizaje (interactúa o recíproco) y 3) regulador del aprendizaje (director o manager).

Otro trabajo muy difundido en la literatura es el de Marsh (1982, 1984), Marsh y Dunkin (1992) y algunos colaboradores. En él se propone como variable dependiente la valoración global del estudiante y de variables independientes nueve factores que son: 1) aprendizaje, 2) entusiasmo, 3) organización, 4) interacción con el grupo, 5) apoyo individual, 6) amplitud y cobertura, 7) justicia en los exámenes, 8) asignaciones y 9) dificultad del curso. Para consolidar y demostrar la solidez de los nueve factores, Marsh y Hocevar (1991) realizan un

¹ Los constructos son: la simulación del interés del profesor, el conocimiento de la materia, habilidad para hablar, naturaleza y valor de los materiales del curso, expansión intelectual del profesor, estimulación del interés, entusiasmo, transmisión intelectual, preparación y organización, claridad y entendimiento, habilidad de elocución, sensibilidad hacia el progreso de la clase, claridad de objetivos, valor de los materiales del curso, justicia e imparcialidad, dirección de clase, feedback de los estudiantes, discusión de clase, cambio intelectual, respeto a los estudiantes, Disponibilidad y ayuda, dificultad y cantidad de trabajo.

análisis factorial confirmatorio al SEEQ² y concluyen en la definición de un constructo que utiliza los nueve factores.

Hemos de destacar que la estructura del SEEQ ha sido considerada y estudiada durante muchos años por un gran número de investigadores y evaluadores en diferentes países del mundo como Estados Unidos, Australia, Hong Kong y España³. Marsh y todos sus colaboradores han creado un cuerpo de literatura muy importante que ha conferido un potente impulso al estudio de la evaluación de la docencia universitaria.

Respecto a los últimos años observamos trabajos como el de Patrick y Smart (1998) donde se identifican tres dimensiones: la habilidad para cambiar al estudiante, organización y presentación. Esta última dimensión está incluida en el estudio de Kwan (1999), aplicado en la Universidad de Hong Kong, en donde se recolectaron 5000 evaluaciones de profesores en 75 departamentos, empleando un instrumento que además de la dimensión organización y presentación, contenía cinco más relacionadas principalmente con: resultado de aprendizaje, motivación, feedback, interacción y ayuda individual.

La interacción del profesor con los estudiantes también está contenida en un instrumento de Ballantyne, Bothwick, y Packer (2000), quienes proponen siete dimensiones para el estudio: organización y preparación de la clase, presentación de la clase, participación de los estudiantes en la clase, aprendizaje de los estudiantes, evaluación de los estudiantes de su aprendizaje, interacción del profesor y los estudiantes, métodos/ áreas de enseñanza y retroalimentación en la enseñanza.

Por medio del análisis factorial, Swartz y otros (1990) identifican dos factores con la eficacia docente: (1) claridad instruccional de presentación, y (2) administración de la conducta de los estudiantes. Lowman y Mathie (1993) también insisten en la utilización de dos factores : (1) intelectualmente estimulante, y (2) relaciones con los estudiantes. No obstante, en otros estudios se identifican

² SEEQ siglas en inglés de Student Evaluation of Educational Quality

³ Marsh, 1983, 1984, 1987, 1991; Marsh y otros, 1985; Marsh y Bailey, 1993; Marsh y Dunkin, 1992; Marsh y Roche, 1992, 1994, 1997, 2000; Marsh y Hocevar, 1991, 1984; Marsh y otros, 1997; Marsh; Touron y Wheeler (1985).

más y diferentes factores de eficacia docente. Por ejemplo, Brown y Watkins (1994) consideran tres factores (1) afectividad, (2) sistematización y (3) estimulación. El mismo número de factores son descritos por Rindermann (2001) los cuales son: estructuración, la competencia docente y entusiasmo. Otros investigadores han sugerido más factores como los siete de Ramsden (1991) o los nueve que destaca Marsh y Dunkin (1992).

De acuerdo con nuestra opinión, los autores mantienen su coincidencia en los factores relacionados con organización, evaluación, interacción, trato con los estudiantes y entusiasmo. Estos continúan siendo los más difundidos y utilizados como criterios de evaluación docente desde la perspectiva de los estudiantes.

En el contexto español, también se observan múltiples trabajos sobre el tema, como los de Alvarez y García (1999), Escudero (1999), Jornet (1988) y Cajide (1994), entre otros.

Entre las diversas propuestas existentes, Tejedor (1990) insiste en siete dimensiones o factores, altamente correlacionadas entre sí, que deben ser empleadas para evaluar al docente: (a) cumplimiento del profesor, (b) calidad y desarrollo del programa, (c) dominio de la asignatura, (d) interacción con los alumnos, (e) recursos utilizados y prácticas, (f) exámenes y (g) valoración global del profesor. A su vez González Such y otros (1999), destacan tipologías de calidad docente como: cumplimiento de obligaciones, conocimiento de la materia, desarrollo de la clase, materiales y programa, actitud del profesor y evaluación.

Un instrumento muy importante de estudiar y de mencionar, por la profundidad en el análisis de sus componentes y por la rigurosidad con que se valida el constructo competencia docente a través del análisis factorial confirmatorio, es el CEDA⁴ de García Ramos (1996, 1997, 1997b, 1998). El CEDA enumera seis dimensiones implicadas en la evaluación de la competencia del docente universitario, que se describen como: 1) Programa – organización de la enseñanza, 2) dominio de contenidos – claridad expositiva, 3) motivación de aprendizaje – incremento del interés del alumno, 4) interacción con el grupo de clase, 5) atención individual al alumno, 6) evaluación – exámenes. Para englobar

⁴ CEDA significa, Cuestionario de Evaluación Docente por el Alumnado

las seis dimensiones mencionadas, el CEDA se divide en tres partes: valoración de la materia, valoración analítica del docente y valoración global del mismo.

Últimamente la búsqueda de las dimensiones para la evaluación de la competencia docente universitaria continúa siendo objeto de estudio y se pueden observar algunos ejemplos destacados en la siguiente tabla.

Nosotros consideramos que de las diferentes propuestas analizadas, muchas de esas dimensiones pueden llegar a ser sintetizadas alrededor de organización, interacción con los estudiantes, preparación, habilidades para comunicar y estimular a los estudiantes y justicia en la nota. En nuestra opinión, las primeras dos dimensiones aparecen con mayor frecuencia en los estudios; tanto en el ámbito internacional como en el español (Marsh, 2000; Ting, 2001; García Ramos, 1997; González Such, 1999; y Tejedor, 1990).

Factores o dimensiones conceptuales de competencia docente universitaria

La competencia docente es el criterio que con mayor frecuencia se utiliza para evaluar al profesorado universitario y desde una perspectiva general, podría decirse es “aquel que demuestra eficacia en el logro de los objetivos que son propios de su trabajo a nivel universitario -docencia, investigación y servicios a la comunidad” (De Miguel, 1998, 72). En otras palabras, lo anterior se refiere según Schalock (1993) a la “preparación específica del docente universitario para cumplir de manera adecuada con las responsabilidades” (p. 67) y las funciones profesionales.

En este sentido, se busca destacar los atributos característicos que un docente debe tener para una práctica segura de su profesión. Tales atributos han de ser una composición interrelacionada a la cultura, el comportamiento y la acción, entre otros. Cuando los profesores logran reunir e integrar dichos componentes, adquieren ciertas competencias que le sirven para enfrentar la enseñanza universitaria con una mejor visión.

Según Alvarez (1999) la definición de la competencia docente es un problema muy complejo e involucra una serie de factores internos y externos del

profesor. No obstante, ello es el resultado de la interacción entre cuatro macro-variables: 1) oficio docente, 2) el arte de ser profesor, 3) las condiciones estructurales y 4) los dilemas y paradojas. Este estudio se fundamenta a partir de la primera macro-variable.

El oficio docente es un conjunto de saberes y concepciones sobre las funciones a desarrollar por el profesor que dan como consecuencia una forma eficaz de enseñar: profesionalidad (dominio de los contenidos, dedicación, preparación de las clases, asunción de roles, innovación docente) conocimiento y características de las necesidades del alumnado (necesidad de acogida, necesidad de motivación, carencia formativas, necesidad de estructuración de estudio), destrezas docentes (claridad expositiva, claridad en los criterios de evaluación, aplicabilidad del contenido de la enseñanza, dominio del proceso de argumentación y demostración, adquisición y utilización de recursos didácticos personales, habilidad para la comunicación con los alumnos, adaptación de la enseñanza). (Alvarez, 1999, 278)

Considerando la propuesta del autor y analizando los estudios sobre la percepción de los estudiantes de sus profesores universitarios, podemos manifestar que los profesores competentes destacan en las evaluaciones en las dimensiones relacionadas con el oficio docente, y que se refieren a: organización, conocimiento de la materia, comunicación, interacción con los estudiantes, entusiasmo y métodos de evaluación.

Organización y conocimiento de la materia

La organización de la materia de clase, es una de las dimensiones más utilizadas en la literatura sobre el tema. Su empleo es común en los diferentes instrumentos propuestos por investigadores, evaluadores y administradores universitarios. Los estudios más prestigiosos de Marsh (2001), Patrick (1998) y Kwan (1999) proponen la utilización de dicho criterio, como uno de los predictores más destacados de competencia docente.

Un profesor organizado es considerado como aquel que planifica con anterioridad su clase, busca ordena los recursos, organiza profundamente todos los componentes, llámese objetivos, actividades, textos y exposiciones, entre otros. García (1997) y Marsh (1987) han demostrado en sus estudios empíricos que esta dimensión tiene una influencia muy fuerte sobre otros criterios de la competencia docente universitaria.

Además, un profesor ha de conocer profundamente lo que enseña, ha de ser un especialista en constante búsqueda de conocimiento, pero se hace necesario combinar los contenidos que involucra su materia, para generar resultados positivos de aprendizaje. Bien es conocido que el conocimiento de la materia no es garantía de éxito; no es suficiente ser un especialista al más alto nivel, es más importante contar con las habilidades que permitan una forma de intercambio con los estudiantes, para que las experiencias de clase tengan algún significado para ellos.

En este sentido, García y Congosto (2000), aseveran que el dominio de contenidos de una asignatura o disciplina es la condición necesaria. Pero, la condición suficiente para que se produzca una acción educativa – formadora eficaz, esta “ligada al resto de procesos que facilitan la comunicación de los contenidos y la motivación de aprendizaje que generan en el estudiante. Ambas condiciones son básicas” (García y Congosto, 2000, p.245).

Podemos manifestar que el dominio de la materia se encuentra muy relacionado con la organización y la planificación de la clase misma. Trent y Cohen (1973), Marsh (2001), De Miguel (1998), Nuhfer (1996), Frey, Leonard y Beatty (1975), Feldman (1997), Marsh y Roche (1993, 1997), Coffey y Gibbs (2001) y Centra (1993), entre otros, han propuesto el empleo de esta dimensión como un factor de eficacia docente.

En los meta-análisis más importantes sobre el tema Cohen (1986, 1987) con un grupo de 41 estudios bien escogidos, el autor encontró un coeficiente de 0.34 entre el rendimiento del estudiante y el conocimiento de la materia, poniendo en evidencia una moderada relación entre ambas variables. No obstante, en el mismo estudio, el autor resalta que se observa una alta correlación, coeficiente de 0.57,

entre la estructura y logro, lo cual está altamente relacionado con la preparación y organización de la clase. Los resultados evidencian el peso de esta dimensión en la evaluación de la docencia universitaria.

Comunicación con los estudiantes

Los profesores deben ser capaces de comunicar sus ideas efectivamente en diferentes formas y también, a diferentes personas. En esta línea, Baugh (1983) y Taylor (1988), expresan que el aprendizaje se facilita cuando el profesor y los estudiantes comparten por medio de la comunicación, significados de ideas o sentimientos con el uso de sonidos convencionales, signos o gestos, apoyados en el empleo de símbolos escritos que son mutuamente entendidos.

La comunicación puede ser aún más efectiva cuando tiene significados culturales que son compartidos por todas las personas en la clase. Esto significa, teóricamente, que dos culturas pueden tener el mismo lenguaje, pero variedad de grados de comprensión de los disímiles significados de las palabras; éstas se pueden escribir igual, pero generalmente tienen diferentes significados que dependen del lugar, el acento, la intensidad y muchos otros elementos más. Collins (1988) y Heath (1983), tratando sobre la comunicación en el aula, asumen que los profesores necesitan reconocer esta realidad, para hacer ajustes que le permitan facilitar el aprendizaje a sus estudiantes.

En una revisión sobre el tema De la Orden (1990) encontró que la habilidad del profesor para implicar eficazmente a los estudiantes en tareas de aprendizaje es un predictor importante de eficacia, pues existen un gran número de estudios correlacionales que lo indican.

Interacción y entusiasmo

Un docente universitario ha de ser capaz de interactuar con los estudiantes, de manera que ayude al crecimiento académico y personal de éstos. Los estudiantes, realizan un mejor aprendizaje en situaciones donde se sienten

cómodos y, de alguna forma, apoyados por los docentes. Cuando los alumnos perciben que existe cierta afinidad y familiaridad, están más atentos y comprometidos con los procesos de clase. Jackson (1999) descarta que el estudiante observa la habilidad del profesor para apoyar a los estudiantes, creando una atmósfera que conduzca al aprendizaje.

Con un proceso adecuado de interacción los estudiantes se sienten libres de expresar su opinión, apoyarse en conversaciones entre ellos, y sobre todo, animan con ello el desarrollo de la clase

Aunque, no podemos creer que los estudiantes quietos, ordenados o atentos se sientan bien dentro de la clase. Los profesores necesitan ser sensibles a las necesidades emocionales de sus estudiantes y responder a sus relaciones interpersonales, buscando diferentes formas de interrelacionarse con ellos, de manera que les permita conocer una parte de sus intereses, necesidades, metas y preocupaciones. Las relaciones interpersonales acercan al profesor a la realidad del estudiante y abren las puertas a la comprensión, la sensibilidad y sobre todo a la humanización del proceso de enseñanza.

Los profesores exitosos tienden a ser los que pueden utilizar un rango de las estrategias de enseñanza y un rango de estilos de interacción, presentándose ante sus estudiantes entusiasmados por los diferentes elementos de la clase. Estos son capaces de aceptar críticas, agilizar los temas difíciles, se muestran atentos y contentos ante sus estudiantes. En fin, estos aspectos se encuentran constantemente en las investigaciones sobre eficacia docente, lo que sugiere según Doyle (1985) que los profesores eficaces ajustan su enseñanza para conocer las necesidades de diversos estudiantes y las demandas que ellos tienen sobre metas, temas y métodos educacionales.

Good (1983) destaca que el uso de diversas estrategias ocurre en el contexto de la "enseñanza activa" que es determinada y diagnóstica, que responde tanto a las necesidades de los estudiantes como del currículo. En este sentido, la interacción juega un papel determinante, ya evidenciado en los estudios de Cohen (1986) y Marsh (2001) quienes encontraron una correlación de 0.52 para la interacción del profesor con el estudiante.

La evaluación de la enseñanza

Los profesores han de promocionar el rendimiento y el conocimiento del estudiante universitario a un nivel más complejo. Foster (1991) asevera que el planeamiento y la evaluación de la enseñanza permiten al docente observar si el estudiante comprende la materia de clase y la relaciona con destrezas que le permiten ascender a niveles más elevados: ésta es la esencia de la docencia y el aprendizaje. El nivel de habilidad, conocimiento y de destrezas, ha de ser evaluado a intervalos regulares para determinar el progreso de los estudiantes y también, para estudiar otras acciones que le permitan al estudiante a superar las áreas o los capítulos más problemáticos de la clase.

No obstante, para lograr una importante relación entre los intereses docentes y estudiantiles Jackson (1999) afirma que debe existir justicia a la hora de valorar a los estudiantes. Los criterios utilizados tienen que ser los mismos para todos los estudiantes sin ninguna distinción que les permita la duda. Los docentes se deben presentar imparciales en sus ponderaciones sobre el rendimiento de los estudiantes. Además, las pruebas han de mantenerse relacionadas con el contenido estudiado en clase y con el tiempo de estudio de los estudiantes.

La dimensión evaluación es muy señalada por los investigadores sobre el tema y es uno de los predictores destacados por los autores más relevantes. García (1997) propone que esta dimensión influye directamente sobre otras dimensiones de evaluación.

Los autores mencionan que los factores mencionados de evaluación del profesor pueden variar de acuerdo al tamaño de la clase, como en el caso de apoyo e interacción. En un meta-análisis muy importante Cohen (1987) reportó los siguientes coeficientes: estructura (0.55), interacción (.52), habilidades (.50), global del curso (.49), total del instructor (.45), aprendizaje (.39), apoyo (.32), evaluación (.30), feedback (.28); interes /motivación (.15) y dificultad (-.04), claridad de los objetivos del curso (0.35), habilidades (0.56), conocimiento sobre la materia (.34) y sensibilidad con el nivel de clase y progreso (0.30). Se observa una

alta correlación entre estructura y logro pero tiene relación con la preparación y organización (.57). Los hallazgos de este autor son muy consistentes para el uso de tales dimensiones en la evaluación de estudiantes de sus profesores universitarios.

ANEXO N° 3 Artículo Concluido

INDAGACIONES TEÓRICAS ACERCA DE LA FIABILIDAD Y VALIDEZ EN LA EVALUACIÓN DOCENTE

Reliability and validity of the student's rating used to evaluate university professor's performance.

Acevedo Alvarez Raziel

Perfil académico y profesional de los autores

Raziel Acevedo Álvarez, doctor en Métodos de Investigación, Diagnóstico y Evaluación Educativa de la Universidad Complutense de Madrid. Profesor catedrático de la Universidad de Costa Rica, Sede de Guanacaste.

Dirección de los autores

Raziel Acevedo Álvarez. Universidad de Costa Rica, Sede de Guanacaste, Liberia. Tel: (506) 680-0461, fax: (506) 690-0607, correo electrónico: razziel.acevedo@ucr.ac.cr

Resumen

El presente artículo, expone una extensa revisión bibliográfica sobre la fiabilidad y validez de los cuestionarios de opinión estudiantil, utilizados para evaluar la competencia docente universitaria, a fin de reunir información sobre el intenso debate, su complejidad, su análisis y sobre su permanente actualidad; porque la docencia se transforma continuamente, los conocimientos y habilidades de hoy se desactualizan mañana. Por ello, es necesario evaluar la actividad permanentemente, pero con instrumentos que cumplan cabalmente estos dos componentes psicométricos. Sin embargo, es importante recalcar que muchas de estas evaluaciones no han sido estudiadas en profundidad y en muchos casos carecen de estos análisis estadísticos; de ahí nace la gran cantidad de inconvenientes y problemas que enfrentan estas evaluaciones. Además, los cuestionarios de opinión han sido contrastados con variables individuales, olvidando que la docencia es un fenómeno multidimensional integrado por un conjunto de elementos contextuales y de esa forma ha de estudiarse, pues algunos de ellos pueden estar asociados con factores ajenos a la docencia universitaria.

Palabras claves: EVALUACIÓN UNIVERSITARIA, EFICACIA DOCENTE, INSTRUMENTOS DE EVALUACIÓN, CUESTIONARIOS, FIABILIDAD, VALIDEZ

Abstract:

Key words: UNIVERSITY EVALUATION, TEACHER COMPETENCE, VALIDITY, RELIABILITY.

The present article, exposes an extensive bibliographical review on the reliability and validity of the questionnaires of student opinion, to evaluate the educational university competition, to bring together information of the intense debate, of its complexity and of the permanent current importance; because the teaching transforms constant, the knowledge and today skills change tomorrow, for this reason, it is necessary to evaluate the activity, with instruments that fulfill these two component psychometrics. Nevertheless, it is important to stress that many of these evaluations have not been studied in depth and lack these analyses, of there is born the great quantity of problems that face these evaluations. In addition, the questionnaires of opinion have been resisted by individual variables, forgetting that the teaching is a multidimensional phenomenon integrated to a set of elements, not isolated and of this form they have to be studied, since some of them can be associated with other factors foreign to the university teaching.

INTRODUCCIÓN:

La universidad como institución de enseñanza superior asume un carácter complejo al tratar de buscar el equilibrio entre la investigación, la acción social y la docencia. Esta última, quizá la más compleja y debatida, porque es un elemento fundamental en el quehacer académico, al constituirse en el área de impacto social responsable de entregar, a las comunidades urbanas y rurales, profesionales en los diversos ámbitos del quehacer humano.

La docencia es vital en la academia, pero en sí misma, es proceso en constante cambio. Día a día, el docente ha de buscar nuevas destrezas y habilidades que le permitan una mejor posición de interacción entre y con sus discípulos. Su formación debe ser permanente y renovada, para enfrentar los retos que impone la enseñanza. En este sentido, baste este elemento como justificación y punto de partida de este estudio, la necesidad de estudiar, conocer y analizar la evaluación docente, pues es un tema de permanente actualidad.

Ahora bien, para aproximarse a la evaluación docente, se requiere de métodos que permitan conocer, valorar e incentivar esta labor, debido a la importancia de contar con herramientas apropiadas para tomar las decisiones oportunas sobre su acción. En este sentido, uno de las técnicas más utilizadas para ello ha sido el cuestionario de opinión estudiantil, porque los (las) estudiantes son la razón de ser de la docencia, por esa razón es vital saber, conocer y entender su pensamiento, su posición frente a la enseñanza que reciben. Ellos (as) son los receptores (as) de todos los conocimientos, actitudes y actividades docentes. Con este elemento encontramos otro argumento que destaca la importancia de nuestro estudio, convirtiéndolo en novedoso día a día, mientras existan estudiantes y docentes, será indispensable estudiar la validez y fiabilidad de los instrumentos utilizados para evaluarlos.

Por otra parte, hablando del uso de los cuestionarios utilizados para evaluar la acción docente, sirven para múltiples decisiones, acarreando numerosas consecuencias en el personal docente, que pueden incentivar o desmotivar su accionar; por ello, es obligatorio contar con instrumentos fiables y válidos para reunir información precisa y exacta sobre el fenómeno, pero, su evaluación debe estar libre de prejuicios e improvisaciones nacidas cuando no existe un proceso riguroso de análisis de sus propiedades psicométricas.

Lamentablemente, aunque se han visto muchos cuestionarios, son pocos los que han sido analizados con la dedicación necesaria para desestimar arbitrariedades e imprecisiones.

Por tanto, este artículo tiene como propósito fundamental profundizar en los estudios realizados sobre la fiabilidad y validez, aplicados en distintas épocas a numerosos cuestionarios de evaluación, utilizados principalmente en Europa, Norteamérica y Australia, a fin de observar sus métodos, sus procesos de análisis y sus conclusiones a fin de mostrar un camino a los interesados en esta actividad; dado que en América Central aún no es común encontrar estos análisis en los cuestionarios de evaluación de las Universidades o entidades de gobierno. Es más, en algunos casos ni siquiera se evalúa la docencia. La realización de una revisión bibliográfica profunda permite mostrar el panorama que ha emergido en cuanto a la confiabilidad y validez de este tipo de evaluación. A partir de este propósito surgen las siguientes interrogantes:

¿Qué significa validez y confiabilidad?

¿Qué métodos se han empleado para medir estas propiedades psicométricas en los instrumentos de evaluación docente?

¿Pueden realmente los (as) profesores (as) confiar en los instrumentos utilizados para su evaluación?

1. LOS CUESTIONARIOS EN LA MEDIDA DE LA COMPETENCIA DOCENTE:

El cuestionario utilizado para conocer la opinión de los (las) estudiantes respecto de la acción docente, es comúnmente empleado de acuerdo con diferentes intereses de los involucrados, llámese administradores (as), profesores (as) y estudiantes (as). Por ejemplo: 1) la administración los tienen en cuenta, como parte del proceso de toma de decisiones relacionadas con la selección de su personal, los incentivos de promoción económicos o académicos y con la permanencia de sus docentes dentro del sistema universitario; 2) a los instructores (as), les suministra información actualizada acerca de la visión que tienen sus estudiantes de la forma y estructura de sus cursos, de los elementos a mejorar y de la calidad de su enseñanza; 3) a los (as) estudiantes les proporciona herramientas que les permite seleccionar los cursos y las personas de su interés; y 4) a los investigadores (as) les es vital, para indagar sobre la eficacia de la enseñanza y el aprendizaje en la universidad. En fin, los resultados aportados por los cuestionarios han sido utilizados para una gran variedad de fines, por tanto es necesario tener la certeza de que estos instrumentos son totalmente fiables y válidos.

Pese a la importancia de los cuestionarios, según Marsh (2003) muy pocos de ellos han sido estudiados profundamente en cuanto a su fiabilidad y validez, es más cuentan con una insuficiente estructura teórica; lo cual incrementa la incertidumbre y el rechazo por parte de los docentes universitarios (as), sobre la calidad, la rigurosidad del instrumento y la relevancia de los datos aportados. En este sentido, la medida de la competencia docente deja una gran cantidad de interrogantes iniciales relacionadas con la fiabilidad y validez.

2.- LA FIABILIDAD, un paso inicial de estudio

La fiabilidad en psicometría se puede conceptualizar de distintas maneras que, a su vez, se traducen en métodos distintos de cálculo, los cuales son totalmente diferentes al proceso de validez⁵.

Cuando hablamos en términos propios de la fiabilidad de estas evaluaciones, nos referimos sobre todo, a dos elementos: *unanimidad* y *estabilidad*:

La unanimidad indica directamente al punto en que los (as) estudiantes son consistentes o *unánimes* en sus juicios cuando diferencian a un (a) docente de otro (a). En otras palabras, pueden observar muy claro cada una de las características de uno u otro profesor (a) y de esa forma emitir un juicio de valor en conjunto dirigido al extremo del cuestionario, sea de forma positiva o negativa su apreciación. Por tanto, su evaluación de la acción docente no sería dispersa, sino que estaría agrupada alrededor de las características docentes.

La estabilidad, esta condición expresa el punto en que las evaluaciones no varían notablemente cuando pasa el tiempo, sino, los (as) estudiantes mantienen los mismos criterios de valoración del docente universitario (a). De esa forma la evaluación de un instructor (a) no cambiará mucho a través del tiempo y sus resultados serán similares con el transcurso de los años.

Estos dos aspectos de la evaluación de la *docencia* universitaria, *unanimidad* y *estabilidad*, son considerados por Overall y Marsh (1980) en un estudio longitudinal muy importante, en donde se explica que la fiabilidad en las *encuestas de estudiantes* se entiende como el *acuerdo relativo* (unanimidad) entre las valoraciones de diferentes estudiantes dentro de la misma clase, bajo la asunción de que cualquier varianza específica del (la) estudiante individual es aleatoria y debería ser considerada como varianza de error. *Es estable* a lo largo de un período de tiempo de varios años, separando los dos conjuntos de estas valoraciones, ello podrá ser incluido como la varianza sistemática cuando los coeficientes de estabilidad a largo plazo se basen sobre sus respuestas individuales.

No obstante, sobre este tema Cruse (1987) señala que los coeficientes de *fiabilidad* indican que los (las) estudiantes puntúan de forma consistente, de la misma manera en ocasiones diferentes a los (as) profesores (as), pero ello no significa que estos evalúan exactamente la docencia universitaria, pues a los estudiantes se les plantea un modelo tradicional de docente, el cual se ha ajustado a ciertos factores característicos del “buen profesor”. Este argumento es un tema debatido por mucho tiempo, principalmente porque los rasgos encontrados por los investigadores, pueden o no, ser apropiados para un tipo de docencia específica, entonces si eso sucede dejaría sin valorar otros rasgos especiales de los (as) docentes universitarios (as). El problema de todo este trabajo radica en que por la multidimensionalidad y complejidad no solo de la labor docente, sino de la misma universidad en su estructura, se hace difícil encontrar un modelo global que ajuste a todas las características involucradas en el campo específico del saber universitario.

⁵ Una exposición detallada sobre los distintos métodos de cálculo puede consultarse en Feldman (1977).

De regreso a nuestro tema de fondo la *fiabilidad* del cuestionario de opinión, se ha determinado a partir de los resultados del análisis de ítems (correlaciones entre respuesta a diferentes ítems diseñados para medir el mismo componente de efectividad docente) y también, por estudios relacionados con las puntuaciones de diferentes estudiantes respecto a la misma clase. Marsh (1987) expresa en los instrumentos bien contruidos, la consistencia interna es normalmente alta, aunque “proporciona una estimación inflada de la fiabilidad, porque ignora la proporción sustancial de error debido a la falta de acuerdo entre diferentes estudiantes y por lo tanto no deberían ser usados en general” (pág. 275); quizá puede ser apropiado utilizarlas para determinar hasta qué punto, las correlaciones entre facetas múltiples se han hecho tan grandes que las facetas separadas no pueden ser distinguidas, como con el procedimiento multirasgo multimétodo. El mismo autor considera que una de las medidas a utilizar, es la fiabilidad de la respuesta media de clase basada en el acuerdo entre los (las) estudiantes de la misma clase o sea la unanimidad.

2.1. La unanimidad

Comúnmente cuando se habla de fiabilidad, se hace referencia a *unanimidad*, que en su concepción más simple, es el grado de acuerdo de los (as) estudiantes al valorar a los (as) docentes en un ítem cualquiera del cuestionario. Por lo tanto, la fiabilidad puede calcularse en cualquier ítem (o en cualquier factor) y no globalmente en todo el cuestionario como si se tratara de un test convencional. Evidentemente lo anterior considera la existencia de fiabilidad, si las diferencias se deben fundamentalmente a que los (as) docentes son distintos (as) y así lo perciben sus estudiantes; no a que los segundos son distintos en su manera de evaluar al profesor. Un ítem o factor será fiable si existe un grado de acuerdo entre los (las) estudiantes al evaluar un profesor (a), no si los ítems ordenan a los (as) docentes de manera semejante.

Villa y Morales (1993) manifiestan que operacionalmente se trata de los coeficientes de correlación denominados intra – clase, relacionados con el análisis de varianza y con el coeficiente de fiabilidad de Cronbach utilizados comúnmente en los test. Estos coeficientes se valoran de 0 a 1⁶ y su magnitud depende fundamentalmente del número de alumnos que responden en cada clase. Una observación importante es que no va a existir fiabilidad detectable si no hay diferencias entre profesores (as): no se puede clasificar bien a los que son semejantes, por eso, estos cálculos requieren disponer por lo menos, de más de treinta estudiantes. Normalmente en estas evaluaciones se integran muchos más. En un estudio muy interesante, relacionado con la fiabilidad y el número de estudiantes, Marsh (1982) encuentra el valor medio de la fiabilidad en torno a 0.90, cuando son unos 25 estudiantes los que responden a los cuestionarios de evaluación y cuando son 10 los que responden, la fiabilidad baja a 0.74, por esa razón la muestra en la evaluación debe superar las 30 personas.

El grado de unanimidad presenta un elevado nivel en cuanto a la consistencia interna de las escalas, entre sujetos, en diferentes momentos del curso y diferentes cursos de

⁶ El valor cercano a uno es considerado como muy bueno y caso contrario en el cercano a cero.

la misma tipología. Lo anterior expone que en la clase predomina la unanimidad de opinión frente a la divergencia, cuando se trata de evaluar a sus profesores (as). En otras palabras, los (las) estudiantes tienen opiniones *unánimes* o similares, en torno a lo que en grupo se consideran la “buena” docencia universitaria y pueden unánimemente dirigirse con similar opinión a aquellos aspectos considerados deficientes del (la) profesor (a) universitario. Esto supone que el mayor grado unanimidad entre los (las) estudiantes determina un buen nivel de fiabilidad en los ítems que componen los cuestionarios de opinión.

2.2. La estabilidad

La *estabilidad* en los cuestionarios de evaluación docente se comprueba a través del tiempo, cuando se evalúa la instrucción con la opinión del mismo estudiante o se evalúa al mismo profesor (a) en dos segmentos separados por meses, semestres u años. Este tipo de estudios se diferencian por el tiempo transcurrido entre las aplicaciones, variando ya sea a mitad y en el final del curso o, en ocasiones, pueden transcurrir meses e incluso años entre estas. Claro está, las condiciones ideales para este tipo de investigación radican en comparar, las puntuaciones dadas por los mismos sujetos, un tiempo después de haber evaluado a un docente determinado. Los diseños longitudinales de investigación y evaluación, han sido una fuerte herramienta para demostrar la estabilidad de las puntuaciones en los cuestionarios de la evaluación docente universitario.

Así mismo, los diseños transversales o transeccionales también han aportado información al respecto, comparando las puntuaciones de estudiantes actuales con un grupo de exalumnos (as) asumiendo que son similares. Según Guthrie (1954) la *evidencia* sobre la estabilidad de la clase para valorar al docente data de 1924, aunque para Albanese (1991) y Hativa (1996) los datos son más recientes. Aunque existan diferencias entre los investigadores en cuanto a la fecha de inicio de estos estudios, lo cierto e importante del caso es que los resultados demuestran estabilidad de las puntuaciones de los (as) estudiantes.

No obstante, una de las críticas hechas a estos estudios, radican en la imposibilidad de decidir quiénes son los evaluadores más importantes, si los antiguos (as) estudiantes o los (as) nuevos (as). La visión de los antiguos (as) estudiantes ofrece una perspectiva que permite diferenciar, si recibieron una enseñanza eficaz cuando más adelante en el tiempo han tenido que llevarla a la práctica; esto podría crear diferencias sistemáticas de juicios con respecto a los que están cursando estudios. El problema de la muestra es quizá el componente más complejo en los estudios longitudinales, pues el anonimato es siempre indispensable. En este sentido, habría que sumar los (as) alumnos (as) repitentes, desertores (as) y los factores involucrados en esas acciones, lo cual integraría un grado mayor de complejidad al estudio.

Sin entrar en la discusión de la muestra de estudio, se puede afirmar que las evaluaciones de los actuales alumnos (as) son importantes, porque son los beneficiarios inmediatos de la enseñanza, son los que en este momento reciben al docente en sus aulas y comparten su enseñanza. Ahora bien, un problema podría surgir dado que su visión de la actividad puede condicionar la eficacia de la enseñanza, por ende, el (la) docente ha de ser lo suficientemente maduro para mantener una estructura cognitiva y actitudinal capaz de enfrentar estos retos. Por otra parte desde nuestra perspectiva, las evaluaciones de los antiguos (as) alumnos (as) son también importantes: tienen más años en la Universidad, a

veces hasta han concluido sus estudios universitarios y han recibido en sus aulas a una gran cantidad de docentes, involucrados en múltiples campos y con diferentes competencias; un hecho relevante el cual les permite contar con una visión más amplia del rol docente, de su competencia, de los aciertos y desaciertos. Por tanto, son capaces de valorar al docente desde una plano mayor.

Ahora bien, sintetizando algunos resultados obtenidos en los estudios sobre la *estabilidad* de la clase, observamos que el cuestionario de evaluación del (la) profesor (a) es estable en considerables períodos de tiempo. Los (as) estudiantes mantienen su opinión del docente aún después de muchos años de haber concluido sus estudios. Carson (1999), en un estudio longitudinal desarrollado entre 1964-1999, advierte que los (las) estudiantes recuerdan muy bien a sus pésimos (as) profesores (as). Marsh y Overall (1981) encontraron una estabilidad media de $r = 0.83$ entre 100 cursos estudiados a intervalo de un año, lo cual es un grado de asociación notablemente alto. McKeachie (1987) examinó que las opiniones de estudiantes (as) correlacionan alto ($r=0.94$) aún después de haber pasado un año o más entre cada aplicación. Como observamos en los estudios internacionales, la estabilidad en los cuestionarios de evaluación docente ha resultado en una correlación positiva muy alta.

Por otra parte Costin, Greenough y Menges (1971) y Gillmore (1973) encuentran correlación, entre 0.79 y 0.87, en las valoraciones de los (as) estudiantes de un mismo instructor (a) y el grado del curso, demostrando con valores muy altos que las evaluaciones no varían mucho entre los (as) estudiantes nuevos (as) y los viejos (as), evidenciando con ello que la discusión de la muestra es irrelevante pues existe un alto grado de asociación entre ambos grupos. En estudios anteriores sobre el tema como el de Drucker y Remmers (1951) hallaron que los (as) alumnos (as) que habían permanecido fuera de la institución por cinco o diez años, promediaban al docente de la misma manera que cuando eran estudiantes regulares de la Universidad. A todas luces parece ser que la tendencia internacional de las investigaciones, demuestran claramente estabilidad entre las medidas de evaluación docente.

Hativa y Raviv (1993) en un estudio similar concluyen que el mismo curso ofrecido por el mismo profesor (a), en diferentes semestres, ofrece índices relativamente estables, a menos que se realice una intervención instruccional específica donde sea implantada una nueva estrategia educativa, pero si mantiene la estructura de enseñanza que ha utilizado durante en todos sus años de labor, la evaluación se conservará estable.

La estabilidad de las respuestas en los cuestionarios ha sido investigada además, para determinar la influencia de las modas en los (as) estudiantes u otros efectos posibles. Las encuestas dadas por el mismo número de educandos en un período corto de tiempo, según Centra (1972), producen resultados altamente estables. Overall y Marsh (1980) recogieron las opiniones medias cinco semanas después de haber concluido el curso y correlacionaban en un 0.77. Un año después, aplicaron nuevamente el instrumento a los mismos (as) estudiantes y de igual forma correlacionaban significativamente. Marsh (1992b) encuentra que hay un perfil de valoración estable a lo largo del tiempo y no muestra cambios sistemáticos, lo cual evidencia la estabilidad mostrada por estas evaluaciones.

De todo este cuerpo científico internacional, se puede concluir que los (as) estudiantes universitarios (as) dicen después (con los años o meses) más o menos lo mismo que decían antes: su opinión sobre el (la) docente se mantiene a través del tiempo con muy poca variación. Aunque quizá se puede pensar que con los años los (as) estudiantes evaluarán de *manera distinta* a los mismos (as) profesores (as), ha sido comprobado en las evaluaciones y retrospectivas realizadas por algunas universidades de habla inglesa, que muchos (as) estudiantes antiguos o graduados continúan evaluando de manera semejante a sus profesores (as), que cuando asistían a cursos regulares; los resultados de estas nuevas evaluaciones mantienen la misma visión de los mejores y los peores profesores (as), de la misma forma que lo hicieron en el pasado.

Evidentemente, no se trata de verdades absolutas y sin matices, pero en la práctica, las evaluaciones longitudinales nos han demostrado, con resultados más o menos satisfactorios, evaluaciones completamente *estables* a lo largo del tiempo, y sugieren, como una perspectiva añadida que no altera las puntuaciones dadas al final de la asignatura. Drucker y Remmers (1950), Centra (1979) y Marsh y Overall (1981) concluyen que las puntuaciones que los (as) profesores (as) reciben, están generalmente correlacionadas en un período de tiempo.

Hemos visto que las valoraciones de estudiantes son *unánimes y estables*, lo que demuestra fiabilidad en aquellos casos reportados internacionalmente, valdría la pena comenzar a revisar instrumentos centroamericanos y sus reportes.

En este devenir, a veces se entiende por fiabilidad la coherencia entre diversos evaluadores, ya sean colegas, directores o alumnos (as), pero esta coherencia es una prueba de validez convergente, no de fiabilidad. Ciertamente *la validez* de este tipo de cuestionarios, es más difícil de comprobar que su fiabilidad, debido la existencia de múltiples indicadores de validez, aunque para Cronbach y Meehl (1955) la validez de constructo reúne todos los diferentes tipos existentes.

3. LA VALIDEZ EN LOS CUESTIONARIOS UNA OPERACIÓN COMPLICADA

La literatura sobre *la validez* de los cuestionarios para evaluar la competencia docente ha originado muy diversas opiniones sobre el tema variando persistentemente unas de otras y logrando confundir más que aclarar. Feldman (1997), Marsh y Dunkin (1992) y Marsh (2001, 1987) demuestran en sus estudios que la validez de los cuestionarios de opinión de estudiantes son multidimensionales, fiables, estables, razonablemente válidos con respecto a una gran variedad de indicadores de eficacia docente. Por el contrario Shadish (1998) y Weinbach (1988) cuestionan la validez indicando que los cuestionarios

presentan muy poca validez o que no tienen del todo, por tanto no deben ser utilizadas para tomar decisiones sobre el empleo de los (as) docentes. Su discurso asevera que, aunque la literatura sobre evaluación docente es extensa, en muchas ocasiones su nivel conceptual y metodológico es muy mediocre. Otros mantienen una posición intermedia como Greenwald y Gillmore (1997) y Franklin y Theall (1996) expresando que las encuestas de opinión de los (las) estudiantes son generalmente fiables y los indicadores de efectividad docente son válidos, aunque es necesaria una evaluación adicional independiente.

3.1. Los años 70 y la nota final

Esta época destaca por una amplia variedad de temas de estudio, sin embargo el que más preocupa es el sesgo producido por la nota final. Snyder y Clare (1976) manifiestan que las valoraciones de estudiantes están sesgada por la nota final del curso. La propuesta expresa, si los (as) profesores (as) asignan buenas notas, los (las) estudiantes valorarán mejor al profesor (a). Refiriéndose a lo anterior Worthington y Wong (1979) advierte, los hallazgos reportados sugieren que la validez del cuestionario de opinión está cuestionada seriamente, cuando un profesor (a) puede recibir mejor valoración al mejorar la nota de sus alumnos (as). Un instrumento de evaluación afectado por esta variable, invalida totalmente la evaluación docente realizada, no se podría considerar para nada la información aportada.

Durante estos años se realizaron una serie de trabajos experimentales, tratando de demostrar el sesgo presentado por las encuestas de estudiantes al introducir la variable de nota final del curso. Según Chacko (1983), Holmes (1972), Powell (1977), Vasta y Sarmiento (1979) y Worthington y Wong (1979) la metodología consistía en manipular la nota del (la) estudiante hacia arriba o hacia abajo, según el objetivo del experimento, para determinar el grado de influencia ejercida sobre la opinión del (la) estudiante al rellenar el cuestionario de evaluación. Indudablemente, estos investigadores encontraron algunos efectos en sus grupos de trabajo y publicaron sus resultados.

No obstante, al contrastar los resultados obtenidos en estos experimentos, utilizando los mismos datos, investigadores (as) de la talla de Abrami, Dickens, Perry y Leventahl (1980); Marsh (1987) y Marsh y Dunkin (1992) encontraron deficiencias en el diseño y en la metodología de investigación, lo cual a su vez invalidaba enfáticamente, los resultados obtenidos. Por tanto, recomiendan no tomarlos en cuenta porque metodológicamente estaban defectuosos. Pese a todo, la línea de los años 70, trabajó mucho sobre la hipótesis de que la nota del curso afectaba los cuestionarios de opinión y fue totalmente apoyada por las investigaciones experimentales realizadas en esos años.

3.2. De los 80 hasta el presente el uso de otras técnicas

En los 80 los estudios e investigaciones trataron de determinar la validez de constructo, suministrando tres tipos de evidencia: los estudios multisección, los path análisis y los estudios multirasgo - multimétodo.

El primero, los estudios *multisección*, es quizás el mejor y mayor grupo de todos los estudios de validez de constructo y se refieren a la multisección de un mismo curso en el cual participaron varios profesores (as), utilizando como criterio de rendimiento un examen cualesquiera aplicado a los (as) estudiantes. Los investigadores correlacionan la media de la valoración del (la) estudiante y la media de logro de un curso Universitario. Una correlación positiva y significativa, era tomada como evidencia de la validez de las encuestas de estudiantes. Los estudios demostraron que las diferencias en el rendimiento de los (las) estudiantes, que fueron atendidos en un mismo curso por diferentes profesores (as), se refleja en las evaluaciones del docente hecha por la clase.

Estos resultados han sido muy analizados en varios meta-análisis, no obstante no se llega a un acuerdo relacionado con la validez de los cuestionarios de opinión, aunque ciertamente se han logrado una modesta validez convergente. Al respecto Abrami, Cohen y d'Apollonia (1988) y finalmente D'Apollonia y Abrami (1997) destacan que las correlaciones entre los exámenes como medidas del logro de los (las) estudiantes tiene una media de $r = 0.40$, valor a tomar con precaución. Marsh y Dunkin (1992) refiriéndose a este índice advierten el resultado puede estar afectado por la variable motivación, en las diferentes secciones o por el grado de satisfacción con la nota y es un elemento a controlar, antes de realizar los análisis de datos de la evaluación docente.

Si revisamos la literatura sobre el tema, podemos ver la existencia de varias publicaciones⁷ sobre los estudios multisección, los cuales han marcado una diferencia clara en cuanto a la validez de los cuestionarios de evaluación. Por un lado Cohen (1981) señala “*existe un fuerte apoyo a la validez de las valoraciones de estudiantes como medidas de la eficacia docente*” (p. 300), por otra parte Dowell y Neal (1982) tienen sus dudas al expresar que “*los estimados de validez de las valoraciones de estudiantes no convencen*” (p.52).

Sin embargo, McCallun (1984) concluye que los resultados de los cuestionarios son homogéneos. Abrami (1990) no apoya esta afirmación y menciona que Rodin y Rodin (1972) encontraron una fuerte y negativa correlación (-0.75) entre las valoraciones y el logro; mientras Centra (1977) reportó una fuerte y positiva correlación (0.92) en sus estudios. De todo este cuerpo de literatura se puede discernir que las conclusiones sobre el método multisección no cuenta con claros resultados en torno a la validez de estas evaluaciones.

Greenwald (1997b) destaca que los estudios de validez multisección a favor de la validez de constructo de las valoraciones de estudiantes, se apoyan en la interpretación de las correlaciones observadas entre las notas y estas valoraciones en términos de efectos paralelos de terceras variables. Para el autor, si la nota correlaciona con la evaluación es fundamentalmente porque los buenos profesores (as) generan altas notas y altas valoraciones, lo cual es válido.

⁷ Ver: Abrami, 1989; Cohen, 1981; 1982, 1983; Dowell y Neal, 1982 y McCallum, 1984.

Otro grupo de estudio son los *Path análisis* y exploran la validez de constructo sobre la idea de una tercera variable, explicando la correlación entre la nota esperada y las evaluaciones, pero considerando terceras variables. Howard y Maxwell (1980) midieron el grado de motivación del (la) estudiante y concluyeron “la relación entre la nota y la satisfacción del (la) estudiante puede ser vista como un resultado de causa importante en la relaciones entre otras variables, por ejemplo, como evidencia de contaminación del grado de indulgencia del profesor” (pág. 810). En otro ejemplo similar Marsh (1980) destaca:

“...los path análisis han demostrado que los (las) estudiantes como sujeto principal de interés tienen un fuerte impacto en las evaluaciones y ... también cuentan como una tercera parte de la relación entre la esperanza de nota y estas evaluaciones...por su parte la expectativa de nota es vista como sesgo – aunque pequeño- en las evaluaciones y esta interpretación está abierta a las interpretaciones alternativas” (p.196).

Con el desarrollo de los modernos métodos de análisis estadístico para estudiar las relaciones causales entre las variables (LISREL, EQS y AMOS), se abre toda una gama de posibilidades para explorar con profundidad la competencia docente universitaria. Ejemplos de este método se pueden observar en los trabajos de García (1996, 1997, 1998) y Acevedo (2004) quienes elaboraron propuestas a partir de los modelos de ecuaciones estructurales en la validez de constructo utilizando Lisrel y Amos respectivamente.

El tercer grupo para estudiar la validez de los cuestionarios de opinión, han sido los estudios *multirasgo - multimétodo*. Esta técnica fue desarrollada por Campbell y Fiske (1959) como un medio para obtener índices acerca de la validez convergente y validez discriminante. El modelo se basa, según López Feal (1986) en la diferencia entre rasgo como atributo, característica o propiedad mensurable y método como forma de aproximación a la medida del rasgo.

Los estudios multirasgo – multimétodo en la investigación sobre la competencia docente se han utilizado para demostrar que la valoración de los (las) estudiantes tiene validez convergente y discriminante. En este sentido los autores como Marsh (1982) y García Ramos (1999a) han tratado de demostrar que las encuestas correlacionan: a) bien con las medidas basadas en otros métodos de evaluación del constructo de la calidad docente y b) relativamente menos bien con medidas de otros constructos. Estos estudios han demostrado la validez convergente y discriminante de los cuestionarios de opinión. No obstante, Marsh (1987) señala que el diseño de estas investigaciones requiere de un gran control por las amenazas a la validez interna, externa y ambas. Específicamente el diseño puede presentar amenazas a la validez interna, pero controlando las características externas se pueden conocer lo que significan las diferencias de clase. Lógicamente estos estudios no se adapta a las características extrañas o ajenas a la actividad docente como las variables de los (las) estudiantes, los cursos y otras.

Por otra parte Greenwald (1997b) destaca que ninguno de los estudios ha considerado la expectativa de nota como un recurso de contaminación en la evaluación de la competencia docente. Acerca de este tema McKeachie (1979) afirma, la mayoría de los factores que pueden invalidar las evaluaciones de estudiantes tienen efectos muy pequeños, los cuales no pueden ser considerados como contaminantes del proceso de evaluación. De la misma forma se expresa Marsh (1984, 1997, 2001) quién manifiesta, posiblemente el grado de indulgencia produzca un efecto de sesgo, pero el apoyo empírico estadístico a esta idea es muy débil y pequeño como para ser tomado en consideración. Las evaluaciones de estudiantes tienden a ser estadísticamente fiables y válidas y relativamente libres de sesgo.

3.3. Nuevas perspectivas de estudio y análisis

Actualmente se toma en cuenta más de un tipo de validez y cada una tiene un fuerte enfoque que marca la diferencia. Por ejemplo Marsh y Roche (1997) centran su trabajo en la estructura conceptual de las valoraciones. Su principal punto de vista es analizar la eficacia docente como un constructo multidimensional. Por tanto, los autores tienden a utilizar la medida de las valoraciones para capturar la amplitud de esas dimensiones. D'Apollonia y Abrami (1997) destacan la validez convergente como el foco principal de su trabajo. De su revisión de la literatura sobre validez multisección, los autores concluyen que las valoraciones presentan correlaciones substanciales entre el logro de los (las) estudiantes y el examen como medida de rendimiento. Greenwald y Gillmore (1997) concentran su trabajo en la validez discriminante, analizando las regularidades observadas entre la correlación de la nota esperada y las valoraciones desde múltiples teorías y perspectivas estadísticas. Los autores concluyen que la más fuerte contribución de la correlación entre la nota y la valoración de estudiantes es imperceptible (aunque estadísticamente correcta). McKeachie (1997) está a favor de la validez convergente y discriminante de las valoraciones de estudiantes pero, sin embargo, su idea es que no deben ser empleadas en muchos grupos.

Un resumen de la posición de los diferentes autores se puede observar en la siguiente tabla.

TABLA N°1 AUTORES Y POSICIONES TEÓRICAS RELACIONADA CON LA VALIDEZ

Autor	Estructura conceptual	Validez convergente	Validez discriminante	Validez consiguiente ⁸
Marsh y Roche (1997)	La eficacia docente es conceptual y empíricamente multidimensional. Su validez y particularmente su uso como feedback son socavados por la ignorancia de su multidimensionalidad.	Las diferentes dimensiones de estas encuestas están consistentemente relacionadas con los criterios conocidos de eficacia docente. Ello apoya la validez de constructo.	Estas valoraciones no se ven afectadas por el sesgo. El sesgo ha sido mal interpretado.	La multidimensionalidad se ve fortalecida por la consulta. La mejora de la eficacia docente es el más importante propósito de las encuestas. Su uso para decisiones personales, podría ser más informativo y sistemático
D'Ap	Aunque la enseñanza es multidimensional las evaluaciones de estudiantes contienen un	Las valoraciones globales de los estudiantes están moderadamente correlacionadas con	Existe una pequeña evidencia de sesgo, aunque pocas de esas características conocidas afectan	Las valoraciones proveen una información válida de la eficacia docente. Sin embargo, no pueden ser el único recurso de información.

⁸ Se refiere al uso que se le da a las evaluaciones y si éstas benefician el sistema educativo.

ollonia y Abrami (1997)	gran factor global.	lo que el profesor produce y el aprendizaje de los estudiantes	las valoraciones.	
Greewald y Gillmore (1997)	Las evaluaciones de estudiantes están dominadas por un factor global evaluativo y muchos de los ítems detectan sólo este factor.	Las medidas de las valoraciones presentan moderada correlación con el logro en los diseños multisección	El mismo instructor puede tener altas valoraciones si brinda altas notas o enseña en clases con pocos alumnos. Los valoraciones aumentan con un estilo entusiasta del (la) profesor (a)	La búsqueda de altas valoraciones está inducida sutilmente por la nota o por la reducción de los contenidos académicos del curso.
McKee achie (1997)	Hay un factor "g" en las valoraciones, pero este es también discriminante de viejos y bajos factores	Las valoraciones de estudiantes son validas, aunque son medidas imperfectas de la eficacia docente	Se ven influenciadas por otras variables que están relacionada con el contexto.	Éstas contribuyen a juzgar la eficacia docente, pero su uso debe estar dirigido hacia la mejora.

Fuente: Greenwald (1997), pág. 1185.

Como se ha visto la investigación sobre la evaluación de estudiantes ha sido conducida en muchas perspectivas. Algunos estudios como los de Barnes y Barnes (1993), Hativa (1996) y Marsh y Overall, (1981) han evaluado su fiabilidad y validez a través del tiempo, de los cursos y de los instructores. Otros estudios han evaluado la validez concurrente de las valoraciones de estudiantes, correlacionándolas con otros supuestos criterios de medida, con el fin de contabilizar la varianza con el constructo multidimensional del rendimiento de clase. Los criterios encontrados que correlacionan substancialmente con las valoraciones de estudiantes de su profesor incluyen: estudiantes antiguos, colegas, evaluadores externos y administradores. Feldman (1989),

halló que el criterio menos correlacionado con estas evaluaciones es el de productividad investigadora del (la) profesor (a) (Noser, 1996) y el logro de los (las) estudiantes (Koon y Murray, 1995). Indudablemente, la explicación a ello podría encontrarse en el hecho de que los (las) estudiantes están menos preocupados por la productividad investigadora de sus docentes y más por la calidad de sus explicaciones, imparcialidad de sus procesos evaluativos, etc.

Más allá de los elementos puramente psicométricos relacionados con la fiabilidad y la validez de este tipo de evaluación, los investigadores han examinado también el nivel y el diseño de estas evaluaciones. Los estudios relacionados con este tema sugieren que la evaluación varía según el tamaño de la clase, el formato de la clase y el nivel de estudios. Broder y Dorfman (1994) estudian el diseño e indican que los mismos estudiantes pueden exhibir diferentes patrones de evaluaciones de un mismo profesor, dependiendo de sus necesidades.

Para Braskamp y Ory (1984) los diversos estudios sobre la validez de los cuestionarios de opinión de los (las) estudiantes de la competencia docente universitaria apuntan hacia *dos rumbos* muy delimitados:

1.- Analizar la medida en que factores extraños pueden *sesgar* los cuestionarios de evaluación.

2.- Estudios correlacionales entre las valoraciones de los (las) estudiantes y otras medidas consideradas indicadores *reales* de la efectividad docente.

Estas dos líneas de investigación han sido objeto de muchos trabajos, crítica y resultados dispares en el mundo de la evaluación docente universitaria, debido a que unos buscan los factores asociados al “buen” docente, a aquellas competencias que se deben tener para manejar con sabiduría la clase. Por otro lado, un cuerpo extenso de investigadores ha tratado de identificar aquellas características externas del estudiante, profesor y la clase que pueden atentar contra la validez de este tipo de evaluaciones.

3.4. Fuentes de sesgo

Respecto a la primera línea de investigación, existen una serie de trabajos relacionados con el contexto de los involucrados y la evaluación de los (las) estudiantes. Wachtel (1998) considera que en los últimos años, el foco de atención de las investigaciones se ha desplazado hacia "*el interés metodológico de las características específicas de contexto que puedan dañar la validez*" (pág.192) de estas evaluaciones. Esto se refiere a la posibilidad de que las características de contexto, factores ajenos a las competencias del (la) profesor (a) universitario (a), y que pueden sesgar las evaluaciones hechas por los (las) estudiantes. De manera semejante expresa Worthington (2002) al considerar que la validez y fiabilidad de estas evaluaciones se ve afectada por las variables del contexto, tanto de los (as) docentes como de los (las) estudiantes. Ting (2001) por su parte, reconoce que existen tres determinantes fundamentales que afectan estas valoraciones: las características contextuales del curso, del (la) estudiante y del (la) profesor (a). Acevedo (2006) propone que estas diferencias se encuentran alrededor de las características del (la) docente y sus pupilos (as).

Relacionado con las *características de contexto* Feldman (1978) Cheng y Hoshower (1998) y Wachtel (1998) han estudiado y analizado las particularidades del proceso de *administración de la evaluación*. Feldman (1977, 1978, 1979) observa que el tiempo, el propósito de la evaluación, el anonimato y la presencia del instructor en clase, pueden posiblemente influenciar los resultados de la evaluación y en alguna manera crear sesgo. Marsh y Dunkin (1992), Braskamp y Ory (1994) postulan que las influencias en la evaluación del (la) docente están relacionadas con las características del curso. En la actualidad, está muy difundido y se reconoce el impacto que tiene de la electividad del curso (si es obligatorio o no), el nivel, el área de conocimiento y la cantidad de trabajo, entre otros elementos.

Anderson y Siegfried (1997) y Wachtel (1998), entre muchos otros⁹, estipulan que la validez de estas evaluaciones tienen relación con las *características del instructor*. Esas características incluyen entre otras: el rango, la experiencia, la reputación, las habilidades de investigación, el género y la apariencia física.

⁹ Ver por ejemplo los textos clásicos de: Centra y Creech, 1976; Feldman, 1983; Hamilton, 1980; Mash, 1980, 1987; Perry, Abrami y Leventhal 1979; Erdle, Murray y Rushton, 1985; Greenwald y Gillmore, 1997; Powell, 1978, 1977; Elmore y Pohlmann, 1978

Koermer y Petelle (1991), Tatro (1995), Cheng y Hoshower (1998) han analizado los factores relacionados con las características de los (las) estudiantes¹⁰ e hipotetizan que el interés del (la) estudiante, el género, su expectativa de nota, su edad tienen una influencia que puede sesgar la evaluación de la competencia docente universitaria.

Cuando examinamos la investigación existente asociada al sesgo en la evaluación de los (las) estudiantes de la docencia universitaria, se observa que existen una gran cantidad de puntos emergentes. Si bien algunas de esas características como la administración de la evaluación, el instructor y el curso han sido extensivamente estudiadas, muy poca atención se ha puesto en determinar cuál es la variable de sesgo más importante, de todas las mencionadas.

Evidentemente, aunque algunos estudios se han empezado a realizar en otros países, como el Reino Unido, España, Australia o Hong Kong, el grueso de las investigaciones realizadas han sido efectuadas exclusivamente en Estados Unidos y Canadá. Un ejemplo de esos trabajos incluye a Tatro (1995), Anderson y Siegfried (1997), Cheng y Hoshower (1998) y Ting (2001). Algunos otros estudios como los de Casey (1997) y Timpson y Anderson (1997) sobre el proceso de evaluación en Australia, no pusieron su atención en el sesgo en estas evaluaciones. Aunque ciertamente pueden existir algunas diferencias significativas entre los (as) docentes de otros países y USA, actualmente no se cuenta con estándares de comparación. A pesar de esto Marsh (1994, 1992, 1988, 1981, 1985, 1997) con la colaboración de otros investigadores, han intentado consolidar la aplicación del instrumento SEEQ, en diferentes países con resultados muy halagadores, lo cual pone en evidencia que existen ciertas características docentes universales.

Por otra parte, la mayoría de los trabajos existentes no han centrado el problema de investigación, en las características del *contexto de la docencia* universitaria y se han enfocado solamente en tomar elementos particulares o micro variables, como si estos no fueran componentes de un conjunto. Este tipo de investigación es un problema porque no muestra la docencia universitaria en toda su complejidad, pues deja al margen de análisis una serie de factores que pueden o no estar enlazados con otros. Aunado a ello, otras variables de contexto como la edad de los (las) estudiantes, el nivel del curso, tamaño de la clase, interés previo de los (las) estudiantes, todavía aún no han sido suficientemente investigadas como un grupo que integra un fenómeno de estudio. Por ello, su estudio aislado no es del todo recomendable, sino más bien utilizando los modelos jerárquicos lineales se podría indagar con mayor profundidad con los que han hecho los estudios de Ting (2002) y Acevedo (2006)

A nuestro juicio consideramos particularmente importante la aplicación de rigurosos métodos de análisis empírico, porque con ello se podrían facilitar con cierta certeza el estatus de las características de contexto en la evaluación de la competencia docente

¹⁰ A manera de ejemplo ver también: Marsh 2001; Zoller, 1992; Perry, 1990; Kember, 1994; Marsh y Ware, 1982; Abrami, Leventhal y Perry, 1982; Braskamp, Ory y Pieper, 1981.

universitaria. En este sentido McKeachie (1997) advierte sobre todo a los investigadores relacionados con el tema de sesgo en estas evaluaciones, que es necesaria mayor precaución en la interpretación de los datos para ofrecer resultados coherentes con la realidad del fenómeno estudiado, porque se están tomando para análisis las variables de manera aislada.

Si bien los estudios de sesgo en estas valoraciones tratan de determinar hasta dónde la diferentes variables afectan a los resultados, otros estudios tratan de determinar más bien la relación positiva con los criterios identificados de competencia docente.

3.5 Correlaciones positivas entre las dimensiones e indicadores de competencia docente

La mayor evidencia sobre la validez de los cuestionarios de opinión de estudiantes, procede de los estudios en los cuales existe correlación positiva en las dimensiones de estos instrumentos, reunidas alrededor de lo considerado como un “buen profesor”, por ejemplo: organización, evaluación, interacción, comunicación y apoyo, entre otras. Además, también se ha observado que *correlacionan positivamente* con otros indicadores como: evaluación por colegas, autoevaluación, valoración por expertos, evaluación por alumnos graduados y por el aprendizaje de los (las) estudiantes. Al respecto Abrami (1990) manifiesta, los investigadores han validado las dimensiones consideradas propias del “buen profesor”, como medidas del proceso instruccional, cuyos hallazgos han determinado que estas dimensiones correlacionan con el juicio total de los (las) estudiantes sobre la competencia docente y con el juicio que hacen los colegas, antiguos estudiantes, administradores y observadores externos¹¹.

Para validar los cuestionarios de opinión se correlacionan con: (1) evaluación de los colegas, (2) juicio de expertos, (3) valoraciones de los (las) estudiantes y graduados, y (4) con medidas de aprendizaje. Todas las asociaciones indican la existencia de una moderada o alta correlación positiva, lo cual viene a ser considerado como una evidencia adicional de la validez de los cuestionarios de opinión de los (las) estudiantes. Todo lo anterior se opone a los estudios realizados por Bendig (1953) y Rodin y Rodin (1972) quienes encontraron correlación negativa entre el logro de los (las) estudiantes y su evaluación del (la) profesor (a). No obstante, escritos posteriores de Centra (1973a), Frey (1973), y Menges (1991) han criticado fuertemente la metodología utilizada por estos investigadores, lo cual ha objetado los resultados obtenidos en los estudios estos.

Por otra parte, en la validación de estas investigaciones se han utilizando herramientas estadísticas tales como análisis correlacionales, análisis factoriales (exploratorios y confirmatorios) y, en menor grado, los modelos jerárquicos lineales. (ver: Aleamori y Hexner, 1980; Burdsal y Bardo, 1986; Greenwald y Gillmore, 1997; Marsh, 1984, 2001; Vandewalle, 1997; Marsh y Roche, 1997; García Ramos, 1996; Tejedor, 1990; Ting, 2001; Goddard, Hoy y Woolfolk, 2001; Acevedo, 2006). Pese a esto

¹¹ Un pequeño ejemplo de los investigadores que han trabajado en esta línea se pueden ver en: Abrami, d'Apollonia y Cohen (1990), Cohen (1989), Dickinson (1990), Drews, Burroughs y Nokovich (1987), Gigliotti y Buchtel (1990), Harrison, Ryan y Moore (1996), Koon y Murray (1995), Nimmer y Stone (1991), O' Connell y Dickenson (1993), Prave y Baril (1993), Prosser y Trigwell (1990), Ryan y Harrinson (1995), y Vu, Marriot, Skeff, Stratos y Litzelman (1997).

McKeachie (1997) destaca que un problema adicional relacionado con la validez de las conclusiones de estas evaluaciones, se fundamenta en el propio uso de los datos debido a la falta de sofisticación estadística en los comités o los encargados de utilizar esta información, parece ser que estos carecen de nivel o formación estadística y pueden expresar o explicar resultados de forma errónea o falsa. Por ello se recomienda contar con personal que tenga un alto grado de especialización en la materia,

Hemos observado la presencia de una moderada correlación del criterio de competencia docente con el logro de los (las) estudiantes, en algunos estudios que investigan esta relación con los predictores de competencia docente, como: d'Appolona y Abrami (1997), Greenwald (1997), Marsh y Roche (1997), McKeachie (1997) y Saroyan y Amundsen (2001). Por otra parte, cuando las evaluaciones se asocian a variables no relacionadas con la eficacia docente, los resultados han sido muy diversos y hasta contradictorios. Mientras algunos estudios, como d'Apollonia y Abrami (1997) y Marsh y Roche (1997) sugieren que existe una pequeña evidencia de sesgo de estas valoraciones, otros, por su parte, como Greenwald y Gillmore (1997) proponen que el grado de benevolencia, el tamaño de la clase y el entusiasmo del instructor son elementos potenciales de sesgo.

4. DISCUSIÓN

El estudio de la competencia del docente universitario es motivo de preocupación desde inicios del siglo XX, pasando por la década de los 50 y se intensifica entre los años 70 y 80. Obviamente, en la actualidad esta área es impostergable debido al rol del (a) docente universitario como actor social, quien debe enseñar a aprender a sus alumnos y a su vez, orientarlo a tomar iniciativas innovadoras en el mundo cambiante del cual es partícipe directo. Indudablemente en los últimos años, la preocupación está centrada en la evaluación institucional como un todo y no en un solo componente como es la acción docente, que es una parte de ella. Sin embargo, la evaluación docente es la de mayor preocupación, mayor estudio e incuestionable actualidad. El hecho de evaluar a los (las) profesores (as) universitarias permanecerá en el tapete de las inquietudes universitarias; conocer la actividad es una necesidad imperiosa en aquellas universidades preocupadas por la calidad de su enseñanza.

La coyuntura sociocultural del presente, exige revisiones científicas sobre la competencia del docente universitario, por ello, la evaluación permanente del docente es labor ineludible, que debe asumirse mediante la aplicación de técnicas e instrumentos analizados rigurosamente, en donde la improvisación y miedo no tengan entrada. Pero para llegar a este nivel, han de utilizarse las más modernas técnicas y métodos de análisis estadístico, el establecimiento de criterios de evaluación consensuados y sobre todo, funciones claras de evaluación. Aunque ciertamente, debemos tener presente los primeros elementos de ésta: ¿Qué evaluar?, ¿Para qué evaluar? ¿Cómo hacerlo? y ¿Con qué propósito se hace?

Respecto al apoyo y la oposición de los implicados en este tipo de evaluación, existen diferentes puntos de vista, que señalan vértices opuestos, dado que aparecen personas a favor y en contra. Aleamori (1981) destaca que en la opinión de los (as) profesores (as) sobre la utilidad de este tipo de evaluación no hay indicios claros pues los (as) docentes se mantienen en los extremos, unos señalan que es fiable, válida y útil y otros

por su parte dicen todo lo contrario. Sin embargo después de casi 70 décadas de investigación en el uso de evaluación de la competencia docente universitaria, se puede manifestar con seguridad que los principales investigadores confían que las encuestas son válidas y fiables y que vale la pena realizarlas. De hecho uno de los más destacados autores Marsh (1984), manifiesta que la evaluación de estudiantes es solamente un indicador de la eficacia docente, que tiene una validez establecida a conciencia y rigurosamente. Pero es solamente uno de los muchos indicadores necesarios para valorar la actividad docente y por lo tanto, es deber de la administración basar sus conclusiones en los numerosos indicadores e instrumentos a fin de conocer realmente el grado de calidad del (la) profesor (a) universitario y no puede apoyar sus conclusiones únicamente en un solo instrumento. Esta recomendación debe permanecer constantemente en el escritorio de los administradores (as) universitarios (as), porque hacerlo de otro modo, sería ignorar todos los procedimientos de evaluación docente que se han utilizado para conocer su actividad en la Universidad.

Claro está que los cuestionarios de evaluación han sido muy utilizados y han sido fuertemente criticados en cuanto a su validez y fiabilidad, sin embargo, son una técnica fiable que ha sido estudiada por muchos investigadores, quienes han reportando coeficientes muy altos. De esa manera lo confirman Feldman (1977, 1984, 1997), quien considera que estos instrumentos son realmente fiables y se han reportado coeficientes de fiabilidad en un rango superior de 0.90¹².

Cabe añadir que si un instrumento no es construido adecuadamente, como usualmente es señalado por los investigadores, obviamente su fiabilidad será muy baja y su validez nula. Por tanto, los resultados de la evaluación serían inútiles, así como los juicios que de ella se deriven. Millman (1981), refiriéndose a lo anterior, destaca que si el instrumento no ha sido adecuadamente construido con la ayuda de profesionales, la fiabilidad será muy baja. De igual forma, Craton y Smith (1990) concluyen que un instrumento bien desarrollado y procesado, puede brindar una fiabilidad interna muy alta, caso contrario, donde quiera que los formularios de valoración de los (las) estudiantes no son cuidadosamente construidos con la ayuda de profesionales, como en el caso de la mayor parte de estas encuestas, la fiabilidad puede ser muy baja.

Ahora bien, hacemos un juicio valorativo al manifestar que toda la evidencia apunta directamente a que las encuestas de opinión de los (las) estudiantes son relativamente fiables, unánimes y estables. Su fiabilidad es más elevada que cualquiera de los otros procedimientos empleados para evaluar la docencia universitaria.

Ahora bien, si el concepto de validez es todavía objeto de mucha controversia y debate, también lo es *la validación* de los cuestionarios de opinión de los (las) estudiantes, dado que no existe un criterio específico sobre lo que es instrucción eficaz. En consecuencia, muchos investigadores utilizan un enfoque de validación criterial, relacionando estas encuestas con otras medidas que se asumen como indicadores de

¹² Por su parte Aleamori (1978a) encuentra también rangos de 0.81 a 0.94 para ítem y de .88 a .98 para las subescalas del CIEQ. A su vez Coffey y Gibbs (2001), en el Reino Unido, y Rindermann y Schofield (2001), en Alemania, obtienen coeficientes de fiabilidad de 0.80 a 0.97. Similares índices han sido localizados en España por investigadores tan relevantes como Tejedor y Montero (1989), Salvador, (1990), Muñiz (1991), Jornet (1995) y Abalde y otros (1995), los cuales han reportado coeficientes de fiabilidad que oscilan entre 0.93 y 0.97, puntaje que es considerado muy alto.

eficacia docente. Desde este enfoque, es necesario que exista relación entre estas y los diferentes indicadores de eficacia docente. Al respecto Hilton (1993) subraya, estas valoraciones son pobres medidas de la eficacia docente y el enfoque de la validación de constructo se ve "*disminuido por la falta de un modelo universal aceptable de "buena enseñanza"*" (p. 567).

Por otra parte, algunos investigadores consideran que las puntuaciones en estas evaluaciones están influenciadas por las metas y las estrategias de enseñanza de los instructores, con lo cual en palabras de Kolitch y Dean (1999), no podrían ajustarse completamente a la concepción de enseñanza y aprendizaje descrita en un instrumento típico de evaluación. Sobre este pensamiento Trigwell y Prosser (1996) demuestran cómo la concepción de enseñanza que se asuma, tiene influencia directa en el tipo de acciones que se ejecuten para desarrollarla. Desde este enfoque, la competencia docente obedecería al pensamiento educativo y a las actividades que el (la) docente desarrolle para el aprendizaje de los (las) estudiantes. McKeachie (1997) cree "*que su eficacia dependería entonces de una definición de sus metas de enseñanza*" (p. 1219) y que "*la mayoría de los formularios de valoración de estudiantes del profesor... focalizan casi completamente la enseñanza convencional de clase*" (p. 1220). D'Apollonia y Abrami (1997) admiten que la definición de eficacia instruccional está ligada a la didáctica tradicional de enseñanza y "*...no necesariamente generaliza a otros contextos instruccionales...*" (p. 1199). De la misma forma se expresa Centra (1993) quién mantiene que "*el típico formulario de opinión de estudiantes es concebido para reflejar la eficacia de la charla, la clase y su discusión, y otros métodos centrados en el profesor*" (p. 47). Al respecto Wilson (1988) argumenta que esos formularios suponen una "*pedagogía conservadora*" que representan siempre a "*un estudiante pasivo y un profesor activo*" (p.90). Indudablemente, estos son elementos a considerar en la elaboración de nuevos instrumentos de evaluación. Obviamente, los existentes que han pasado por el proceso de validación han comprobado constantemente su validez, la cual ha sido redundantemente identificada, comentada y alcanzada por medio de múltiples métodos y en diferentes continentes. Marsh (1987), Marsh y Roche (1994) y Ramsden (1991) en sus estudios deducen que la evaluación de los (las) estudiantes es más útil, precisa y válida que otras medidas del rendimiento docente y tiene beneficios añadidos por ser una medida directa de la satisfacción del consumidor. Fundamentado en sus investigaciones, Marsh (1987) concluye que los SET's son probablemente "*el más grande estudio de todas las formas de evaluación personal, y una de las mejores en términos de apoyo para la investigación empírica*". (p. 369)

En síntesis podemos retomar algunas advertencias que han sido destacadas en distintos puntos de nuestra exposición y creemos necesario retomarlas porque nos brindan una visión interrogante que debe ser constantemente replanteada.

La idea de que las dimensiones de evaluación docente presumen ser *un consenso* no existe. Es obvio, ¿cómo podemos evaluar la eficacia docente adecuadamente si no estamos de acuerdo en lo que constituye la eficacia docente?.

La enseñanza es *un arte* y un sentimiento, involucra fomentar cualidades similares, las cuales no son fácilmente evaluables por instrumentos de evaluación.

Los (as) profesores (as) pueden sentir la *pérdida del tiempo* de clase en la administración de los formularios de evaluación y eso les puede desmotivar para experimentar con sus métodos de enseñanza, es importante revisar el tiempo de aplicación. (Centra 1993, p.93)

Los (as) profesores (as) y administradores (as) tienen *poco conocimiento* de la investigación existente en el área y por tanto suelen administrar la evaluación indebidamente. (Franklin y Theall, 1989)

Los instrumentos bien contruidos se ven poco o casi nada afectados, por características ajenas a la competencia docente, son totalmente válidos y podemos confiar plenamente en ellos. Por el contrario, muchos cuestionarios están mal elaborados y eso es un problema constante en las instituciones.

Los resultados de la valoración de la eficacia del docente, constituyen un reto para las autoridades universitarias, quienes deben de velar por la inserción de programas de actualización y mejoramiento de las competencias de los (as) docentes universitarios en pro de garantizar profesionales competitivos acordes a las demandas de la sociedad actual y en respuesta al contexto que caracteriza la universidad como semillero de vocaciones profesionales en diversos campos.

5. A MANERA DE CONCLUSIÓN Y CIERRE

En respuesta a las interrogantes planteadas en el presente artículo se puede decir:

1) La confiabilidad trata de los coeficientes de correlación denominados intra – clase, relacionados con el análisis de varianza y con el coeficiente de fiabilidad Alfa de Cronbach utilizados comúnmente en los test. Estos coeficientes se valoran de 0 a 1, siendo el mejor valor el que está cercano a la unidad. Una gran cantidad de estudios dividen la muestra en dos mitades y le estiman el coeficiente de fiabilidad Alfa, para ver si obtienen los mismos resultados que con la muestra total.

2) Hablar de validez es más complicado, debido a su relación con una estructura teórica, pues ésta se refiere a si realmente estoy midiendo lo que pretendo medir y no estoy midiendo otra cosa. Ello tiene que partir de una sólida estructura teórica la cual, en el caso de la evaluación docente ya existe una gran aproximación al profesor (a) que queremos.

3) Metodológicamente hablando se han utilizado diversas técnicas estadísticas para estudiar la validez de estos cuestionarios, por ejemplo: correlaciones entre factores, estudios multisección, multirasgo – multimétodo, modelos de ecuaciones estructurales (path analysis) y modelos jerárquicos lineales. Estos dos últimos son los más sofisticados y profundos, permitiéndonos indagar con mayor precisión que otras técnicas análisis de datos. En los modelos de ecuaciones estructurales encontramos paquetes como: Lisrel, EQS y Amos. En los jerárquicos: HML y MlwiN, todos cuentan con herramientas muy profundas de trabajo que superan fuertemente los estudios estadísticos tradicionales.

4) Desde nuestra visión, la fiabilidad y la validez de los cuestionarios de opinión ha sido ampliamente demostrada, obviamente, si un instrumento de evaluación docente está mal elaborado, aportará datos erróneos e inválidos y podrá ser cuestionado por los (as) docentes. Por tanto, lo importante será buscar especialistas y fundamentos teóricos que permitan desarrollar un instrumento robusto y consistente.

5) Los (as) docentes universitarios pueden confiar en los resultados de la valoración de su desempeño docente, siempre y cuando apliquen cuestionarios debidamente confeccionados por especialistas y con demostrados índices de fiabilidad y validez.

6) Finalmente, no todos los cuestionarios que se han utilizado en el transcurso de la historia, cumplen con los índices de calidad y ajuste estadístico como para ser utilizados en la evaluación de los (as) docentes universitarios. En este sentido, las autoridades

universitarias deben velar por la confección objetiva de instrumentos de evaluación, que permitan obtener resultados asertivos para la toma de decisiones, ya sea en la contratación o ascenso de profesionales al servicio de la docencia en este nivel de formación.

6. REFERENCIAS

Abalde, E.; De Salvador, X; González Carbanach, R. Y Muñoz C.; J.M. (1995). Análisis de la Evaluación de la Docencia Universitaria por los (as) alumnos (as) en la Universidad de la Coruña (1993-1994). **En Estudios de Investigación Educativa en Intervención Pedagógica**, pp. (289-292) Valencia: AIDIPE.

Abrami, P. C., Dickens, W. J., Perry, R. P., & Leventhal, L. (1980). Do teacher standards for assigning grades affect student evaluations of instruction. **Journal of Educational Psychology**, **72**, 107-118.

Abrami, P.C. (1989a). How Should We Use Student Ratings to Evaluate Teaching?. **Research in Higher Education**, **30** (2), 221-227.

Abrami, P.C.; Cohen, P.A. & D'apollonia, S. (1988). Implementation Problems in Meta- Analysis. **Review of Educational Research**, **58**, pp. 151-179.

Abrami, P.C.; D'apollonia, S. & Cohen, P. (1990). .Validity of Student Ratings of Instruction. What We Know and What We Do Not. **Journal of Educational Psychology**, **82** (2), pp.219-231.

Abrami.C. y D'apollonia, S. (1990b). .The dimensionality of ratings and their use in personnel decisions. M. Theall Y J. Franklins (Eds.), **In Student Ratings of Instruction. Issues for Improving Practice** (pp. 97-111). New Directions for Teaching and Learning.

Abrami, P.S.; D'apollonia, S. & Rosenfield. (1997). .The Dimensionality Of Student Ratings of Instruction. What We Know and What We do Not.. R.P. PRRY & J.C. SMART (Eds.) **In Effective Teaching in Higher Education. Research and Practice**, pp.321-367. New York. Agathon Press.

Acevedo Alvarez, R. & Rodríguez, N. M. (2006). Factores de sesgo asociados a la validez de la evaluación docente universitaria: un modelo jerárquico lineal. **Archivos Analíticos de Políticas Educativas**, **14** (34). Recuperado [27 marzo 2007] de <http://epaa.asu.edu/epaa/v14n34/>

Acevedo, R. y Fernández, M. J. (2004): "La percepción de los estudiantes universitarios en la medida de la competencia docente: validación de una escala". **En Revista de Educación**. N° 28 (2): 145-166. San José: Editorial de la Universidad de Costa Rica.

Albanese, M.A. (1991). .The validity of lecturer ratings by student and trained observers. **Academic Medicine**, **66** (1), pp. 26-28.

Aleamori, L. (1978). .Development and factorial validation of the Arizona Course/ Instructor Evaluation Questionnaire. **Educational and Psychological Measurement**, **38**, 1063-1067.

Aleamori, L. (1981). .Student ratings of instruction.. In J. MILLAMN (ed.), **Handbook of Teacher Evaluation**, pp.110-145. Newbury Park, CA. Sage.

Aleamori, L.M. (1999). .Student Rating myths versus research facts from 1924 to 1998.. **Journal of Personnel Evaluation in Education**, **13** (2), pp. 153-166.

Aleamori, L.M. y Hexner, P.Z. (1980). .A review of the research on student evaluation and a report on the effect of different sets of instructions on student course and instructor evaluation. **Instructional Science**, **9**, pp. 67-84.

Aleamori, L.M. y Yimer, M. (1974). Graduating Senior Ratings' Relationship to Colleague Rating Student Rating. **Research Productivity and Academic Rank in Rating Instructional Effectiveness (Research Report N°352)**. Urbana. University of Illinois, Office of Instructional Resources, Measurement and Research Division.

Anderson, K.H. y Siegfried, J.J. (1997). Gender differences in rating the teaching economics. **Eastern Economic Journal**, **23** (3), pp. 347-357.

Arubayi, E. (1987). .Improvement of Instruction and Teacher Effectiveness. Are Student Ratings Reliable and Valid?. **Higher Education**, **16**, pp. 267-288.

Barnes, L.B. y Barnes, M.W. (1993). .Academic discipline and generalizability of student evaluations of instruction. **Research in Higher Education**, **34** (2), pp. 135-149.

Bendig, A.W. (1953). .Relation of level of course achievement of student, instructor and course ratings in introductory psychology. **Educational and Psychological Measurement**, **13**, pp. 437-488.

Braskamp, L.A. & Ory, J.C. (1994). *Assessing Faculty Work. Enhancing Individual and Institutional Performance*. San Francisco. Jossey-Bass.

Braskamp, L.A.; Ory, J.C. & Pieper, D.M. (1981). Student Written Comments. Dimensions of Instructional Quality. **Journal of Educational Psychology**, **73** (1), pp. 75-70.

Broder, J.M. & Dorfman, J.H. (1994). Determinants of Teaching Quality; What's Important to Students?. **Research in Higher Education**, **35** (2), pp. 235-249.

Burdal, C.A. & Bardo, J.W. (1986). Measuring Student's Perceptions of Teaching. Dimensions of Evaluation. **Educational And Psychological Measurement**, **56**, pp. 63-79.

Campbell, D.T. y Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait – multimethod matrix. **Psychological Bulletin**, **56**, pp. 81-105.

Carson, B.H. (1999). Bad News in the Service of Good Teaching. Students Remember Ineffective Professors. **Journal On Excellence In College Teaching**, **19** (1), pp.91-105.

Casey, R.J.; Gentile, P. y Bigger, S.W. (1997). Teaching appraisal in higher education. an Australian perspective. **Higher Education**, **34** (3), 459-482.

Cashin, W. E. (1995). **Student Ratings Of Teaching. The Research Revisited**. IDEA Paper No. 32. Manhattan, KS: Kansas State University, Center for Faculty Evaluation and Development.

Cashin, W.E. (1988). **Student Ratings of teaching. A summary of the research**. IDEA. Paper No 20. Manhattan, KS. Kansas State University, Center Faculty Evaluation and Development.

CHACKO, T.I. (1983): "Student ratings of instruction: a function of grading standards". En **Educational Research Quarterly**, **83** (1), pp. 19-25.

Centra, J. (1977). Student Ratings of Instruction and Their Relationship to Student Learning. **American Educational Research Journal**, **14**, pp. 17-24.

Centra, J.A. (1972). **The Utility of Student Ratings for Instructional Improvement**. Princton, NJ. Educational Testing Services.

Centra, J.A. (1979). **Determining Faculty Effectiveness**. San Franciso: Jossey – Bass.

Centra, J.A. (1993). **Reflective Faculty Evaluation. nancing Teaching and Determining Faculty Effectiveness**. San Francisco: Jossey-Bass.

Cheng, Y. y Hoshower, L.B. (1998). Assessing student motivations to participate Teaching evaluations. an application of expectancy theory. **Issues in Accounting Education, 13** (3), pp. 531-549.

Coffey, M. & Gibbs, G. (2001). The Evaluation of the Student Evaluation of Educational Quality Questionarie (SEEG) in U.K. Higher Education. **Assessment & Evaluation in Higher Education, Vol 26**, (1), pp. 89-93.

Cohen, P. A. (1981). Student ratings of instruction and student achievement. A meta-analysis of multisection validity studies. **Review of Educational Research, 51**, 281-309.

Cohen, P.A. (1983). A selective review of the validity of student ratings of teaching. **Journal of Higher Education, 54**, 448-458.

Cohen, P.A. (1989). Do grades influence students' evaluation of clinical courses?. **Journal of Dental Education, 53** (4), pp. 238-240.

Costin, F. (1968). **Survey of Opinions About Lectures**. University of Illinois: Department of Psychology.

Costin, F.; Greenough, W.T. y Menges, R.J. (1971). Student Ratings Of College Teaching. Reliability, Validity and Usefulness. **Review Of Educational Research, 41**, 511-535.

Craton, P. y Smith, R. A. (1990). Reconsidering the Unit of Analysis. A Model of Student Ratings of Instruction. **Journal of Educational Psychology, 82** (2) pp. 207-212.

Cruse, D. B. (1987). Student Evaluation and the University Professor. **Higher Education, 15** (6), pp. 723-737.

D'apollonia, S. & Abrami, P. C. (1997). Navigating Student Ratings of Instruction. **American Psychologist**, **51** (11), pp. 1198-1208

De Neve, H. M. F. y Janssen, P. J. (1982). Validity of Student Evaluation of Instruction. **Higher Education**, **11** (5), pp. 543-552.

Dickinson, D. J. (1990). The relationship between ratings of teacher performance and student learning. **Contemporary Educational Psychology**, **15**, pp. 142-151.

Dowell, D. A. , & Neal, J. A. (1982). A selective view of the validity of student ratings of teaching. **Journal of Higher Education**, **53**, 51-62.

Doyle, K. O. (1975). **Student Evaluation of Instruction**. Lexington, MA: Lexington Books.

Drews, D. R. ; Burroughs, W. J. Y Nokovich, D. A. (1987). Teacher self ratings as a validity criterion for student evaluation. **Teaching of Psychology**, **14** (1), pp. 129-143.

Drucker, A. J. Y Remmers, H. H. (1950). Do Alumni and Students Differ in Their Attitudes Toward Instructors?. **Purdue University Studies in Higher Education**, **70**, pp. 62-64.

Drucker, A. J. Y Remmers, H. H. (1951). Do Alumni and Students Differ in Their Attitudes Toward Instructors?. **Journal of Educational Psychology**, **42**, pp. 129-143.

Erdle, S.; Murray, H.G. & Rushton, J.P (1985): "Personality, Classroom Behavior and Student Ratings of College Teaching Effectiveness: A Path Analysis". En **Journal of Educational Psychology**, **77** (4), pp. 394-407.

Feldman, K. A. (1977). Consistency and Variability among College Students in Rating Their Teachers and Courses. **Research in Higher Education**, **6** (2), pp. 223-274.

Feldman, K. A. (1978). Course characteristics and college students' ratings of their teachers. What are know and what we don't know. **Research in Higher Education**, **9** (2), pp. 199-242.

Feldman, K. A. (1979). The Significance of Circumstances for College Students ratings of Their Teachers and Courses. **Research in Higher Education**, **10** (2), pp. 149-172.

Feldman, K. A. (1984). Class size and college students' evaluation of teachers and courses. a closer look. **Research in Higher Education**, **21** (11), pp. 45-116.

Feldman, K. A. (1989a). Instructional effectiveness of college teachers as judged by teachers themselves, current and former student, colleagues, administrators and external (neutral) observers. **Research in Higher Education**, **30** (2), pp. 137-194.

Feldman, K. A. (1989b). The association between student ratings of specific instructional dimensions and student achievement. refining and extending the synthesis of data from multisection validity studies. **Research in Higher Education**, **30**, pp. 583-645.

Feldman, K. A. (1997). Identifying Exemplary Teachers and Teaching. Evidence from Student Ratings. R. P. En PERRY & J. C. SMART (eds.), **Effective teaching in Higher Education. Research and Practice**, . (pp. 368-395). Bronx, N. Y: Agathon.

Fernández, J. ; Mateo, M. & Muñiz, J. (1998). Is There Relationship Between Class Size and Student Ratings of Teacher Quality?. **Educational And Psychological Measurement**, **58** (August), 596-604

Franklin, J. & Theall, M. (1989). **Who read ratings. Knowledge, attitude, and practice of users of student ratings of instruction**. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Franklin, J. & Theall, M. (1989). **Who read ratings. Knowledge, attitude, and practice of users of student ratings of instruction**. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Frey, P. W. (1973). Student Ratings of Teaching. Validity of Several Rating Factors. **Science**, **182**, pp. 83-85

García Ramos, J. M. (1999^a). Análisis multirasgo- multimétodo en la validación de instrumentos de medida para la evaluación de la calidad docente en instituciones universitarias. **Revista Española de Pedagogía**, **214**, pp. 417-444.

García Ramos, J. M. (1997). Valoración De La Competencia Docente Del Profesor Universitario. Una Aproximación Empírica. **Revista Complutense De Educación**, **8** (2), pp. 81-108

García Ramos, J. M. y Congosto Luna, E. (1996). **Un Modelo de Evaluación Institucional en la Universidad**. Salamanca: Studia Pedagógica

Gigliotti, R. Y Buchtel, F. (1990). Attributional Bias and Course Evaluations . **Journal of Educational Psychology**, **82** (2), pp. 341-351.

Gillmore, G. (1973). **Estimates of Reliability Coefficients for Items and Subscales of the Illinois Courses Evaluation Questionnaire**. (Research Report N°341). Urbana: University of Illinois, Office of Instructional Resources, Measurement, and Research Division.

Gillmore, G. M. ; Kane, M. T. & Maccarato, R. W. (1978). The Generalizability of Student Ratings of Instruction. Estimation of the Teacher And Course Components. **Journal of Educational Measurement**, **15**, pp. 1-13.

Gilmore, G. (1984). Student ratings as a factor in faculty employment decisions and periodic review. **Journal of College and University Law**, **10**, 557-576 .

Goddard, R; Hoy, W & Woolfolk, A. (2001). Collective Teacher Efficacy. Its Meaning, Measure, and Impact on Student Archievement. **Collective Teacher Efficacy**. Manuscrito sometido para publicación. 1-40.

Goldman, L. (1993). On erosion of education and the eroding foundation of teacher education (or why we should nor take student evaluation of faculty seriously). **Teacher Quarterly**, **20** (2), pp. 57-64.

Greenwald, A. G. (1997b). Validity Concern and Usefulness of Student Ratings Of Instruction. **American Psychologist** , **51** (11), 1182-1186.

Greenwald, A. G. Y Gillmore, G. M. (1997). Grading Leniency is a Removable Contaminant of Student Ratings. **American Psychologist** , **51** (11), 1209-1217.

Guthrie, E. R. (1954). **The Evaluation of Teaching**. A Progress Report. Seatle: University of Washington.

Harrison, P. D. ; Ryan, J. M. Y Moore, P. S. (1996). College student's self-insight and common implicit theories in the ratings of teaching effectiveness. **Journal of Educational Psychology**, **88** (4), pp. 775-782.

Hativa, N. (1996). University instructors' rating profiles. Stability over time, and disciplinary differences. **Research in Higher Education**, **37** (3), pp. 341-365.

Hativa, N. y Raviv, A. (1993). Using a single score for summative teacher evaluation by students. **Research in Higher Education**, **34** (5), pp. 625-646.

Hepworth, D. Y Oviatt, B. E. (1985). Using student course evaluations. findings, issues and recommendations. **Journal of Social Work Education**, **21** (3), pp. 105-112.

Hilton, H. (1993). Realibility and Validity of Student Evaluation. Testing Models versus Survey Research Model, p. s. **Political Science & Politics**, **26**, pp. 562-569.

Hogan, T. P. (1973). Similarity of student ratings across instructors, courses and time. **Research in Higher Education**, **1**, pp. 149-154.

Holmes, D. S. (1972). Effects of grades and disconfirmed grade expectancies on students' evaluations of their instructor. **Journal of Educational Psychology**, **63**, 130-133.

Howard, G. S. , Conway, C. G, & Maxwell, S. E. (1985). Construct validity of measures of college teaching effectiveness. **Journal of Educational Psychology**, **77**, 187-196.

Howard, G. S. Y Maxwell, S. E. (1980). Correlation between student satisfaction and grades. a case of mistaken causation?. **Journal of Educational Philosophy**, **72** (4), pp. 810-820.

Howell, A. J. Y Symbaluk, D. (2001). Published Student Ratings of Instruction. Revealing and Reconciling the Views of Student and Faculty. **Journal of Educational Psychology**, **(4)**, 790-796

Jornet, J. M. ; González Such, J. Y Pérez Carbonell, A. (1995) . Evaluación De la Actividad Universitaria. G. Quintás (eds.), En **Reforma y Evaluación de la Universidad**. Valencia: Servei de Publicacions de la Universitat de Valencia. pp. 189-244.

Kikut, Lorena (2003). **Exploración de la confiabilidad y validez del cuestionario de evaluación docente**. San José : Universidad de Costa Rica.

Koblitz, . N. (1990). Are student ratings unfair to women?. **Newsletter of the Association for Women in Mathematics, 20**, pp. 17-19.

Koermer, C. D. Y Petelle, J. L. (1991). Expectancy violation and student rating of instruction. **Communication Quarterly, 39** (4), pp. 341-350.

Kolitch, E. & Dean, A. V. (1999). Student Ratings of Instruction en the U. S. A. Hidden Assumptions and Missing Conceptions About . Good. Teaching. **Studies in Higher Education, Vol. 24**, (1), pp. 27-42.

Koon, J. Y Murray, H. G. (1995). Using multiple outcomes to validate student ratings of overall teacher effectiveness. **Journal of Higher Education. 66** (1), pp. 61-81.

Lin, W. Y. ; Watkins, D. Y Meng, Q. M. (1995). Student's Evaluation of University teaching. A China perspective. **Higher Education and Research y Development, 14** (1), pp. 61-74.

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

Mahmound, M. M. (1991). Descriptive models of student decision behaviour in evaluation of higher education. **Assessment and Evaluation in Higher Education, 16** (2), pp. 133-148.

Marques, T. E. ; Lane, D. M. Y Dorfman, P. W. (1979). Toward the development of a sistem for instructional evaluation. Is there consensus regarding what constitutes effective teaching?. **Journal of Educational Psychology, 71**, pp. 840-849.

Marsh, H. & Roche, L. R. (1997). Making Students' Evaluation of Teaching Effectiveness Effective. The Critical Issues of Validity, Bias, and Utility. **American Psychologist, Vol. 52**, 11, 1187-1197.

Marsh, H. (1982). SEEQ. a Reliable, Valid, and Useful Instrument for Collecting Students Evaluation of University Teaching. **British Journal of Psychology, 52**, pp. 77-95.

Marsh H. (1984). Student's evaluation of university teaching; dimensionality, reliability, validity, potential biases and utility. **Journal of Educational Psychology**, 76 (5), 707-754.

Marsh, H. (1987a). Students' evaluation of university teaching; Research findings, methodological issues, and directions for future research. **International Journal of Educational Research**, 11, pp. 253-288.

Marsh, H. (1987b). Student Evaluations of Teaching. M. J. DUNKINS (eds.), **The International Encyclopedia of Teaching and Teacher Evaluation**. (pp. 181-187) Oxford: Pergamon Press

Marsh, H. y Bailey, M. (1993). Multidimensional Students' Evaluations of Teaching Effectiveness. **Journal of Higher Education**, 64 (1), pp. 1-18.

Marsh, H. A. y Overall, J. U. (1979b). Validity of student's evaluation of teaching. A comparison with self evaluations by teaching assistants, undergraduate faculty, and graduate faculty. Paper presented Annual Meeting of the American Educational Research Association, San Francisco.

Marsh, H. W. (1977). The validity of students' evaluations of instructors independently nominated as best and worst teacher by graduating senior. **American Educational Research Journal**, 14, pp. 441-447.

Marsh, H. W. (1980). The influence of student, course, and instructor characteristics on evaluations of university teaching. **American Educational Research Journal**, 17, 219-237.

Marsh, H. W. (1992b). **A Longitudinal Perspective of Student's Evaluations of University Teaching. Ratings of The Same Teacher over a 13 Year Period**. Documento presentado en Annual Meeting of the American Educational Research Association (p. 18) San Francisco, Ca. Abril.

Marsh, H. W. (2001). Distinguishing between good (useful) and bad workloads on Students' Evaluations of teaching. **American Educational Research**, 38 (1), pp. 183-212.

Marsh, H. W. y Dunkin, M. J. (1992). Students' Evaluation of University Teaching. A Multidimensional Perspective. J. SMART (ed.) **Higher Education. Handbook of Theory and Research**. (pp. 143-223) New York: Agathon.

Marsh, H. W. y Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept. First and higher order factor models and their invariance across groups. **Psychological Bulletin**, **97**, pp. 562-582.

Marsh, H. W. y Overall, J. U. (1981). The relative influence of course level, course type, and instructor on students' evaluations of college teaching. **American Educational Research**, **18**, pp. 103-112.

Marsh, H. W. y Roche, L. (1992). **The Use of Student's Evaluations of University Teaching To Improve Teaching Effectiveness**. Canberra: Australian Government Publishing Services.

Marsh, H. W. y Roche, L. A. (1994). The Use of Student Evaluations of University Teaching in Different Settings. The Applicability Paradigm. **Australian Journal of Education**, **36** (3), pp. 278-300.

Marsh, H. W. y Roche, L. A. (2000). Effects of grading leniency and low workload on students' evaluation of teaching. popular myth, bias, validity or innocent bystanders?. **Journal of Educational Psychology**, **92** (1), pp. 202-228.

Marsh, H. W. ; Balla, J. R. y Mcdonald, R. P. (1988). Goodness of Fit Indices in Confirmatory Factor Analysis. The Effects of Sample Size. **Psychological Bulletin**, **103**, pp. 391-410.

Marsh, H. W. ; Hau, K. T. ; Chung, C. M. & Siu, T. L. P. (1997). Students' s Evaluations of University Teaching. Chinese Version of The Student's Evaluations of Educational Quality Instrument. **Journal of Educational Psychology**, **89** (3), pp. 568-572.

Marsh, H. & Roche, L. R. (1997b). Making Students's Evaluation of Teaching Effectiveness Effective. The Critical Issues of Validity, Bias, and Utility. **American Psychologist**, **Vol. 52** (11), pp. 1187-1197.

Marsh, H. W. ; Touron, J. y Wheeler, B. (1985). Students' s Evaluations of University Instructor. The Applicability of American Instrument in a Spanish Setting. **Teaching and Teacher Education**, **1**, pp. 123-138.

Mccallum, L. W. (1984). A meta-analysis of course evaluation data and its use in the tenure decision. **Research in Higher Education**, **21**, 150-158.

McKeachie, W. J. (1979). Student ratings of faculty. A reprise. **Academe**, **65**, pp. 384-397.

McKeachie, W. J. (1987). Instructional evaluation. current issues and possible improvement. **Journal of Higher Education**, **58** (3), pp. 344-350.

McKeachie, W. J. (1990). Research on College Teaching. The Historical Background. **Journal of Educational Psychology**, **82** (2), pp. 189-200.

McKeachie, W. J. (1997). Student Ratings. The Validity of Use. **American Psychologist**, **52** (11), pp. 1218-1225.

McKeachie, W. J. ; LIN, Y. G. y MENDELSON, C. N. (1978). A small study assessing teacher effectiveness. Does learning last?. **Contemporary Educational Psychology**, **3**, pp. 352-357.

Meeth, L. R. (1976). The stateless art of teaching evaluation. report on teaching. Vol. 2, Change, 8, pp. 3-5.

Menges, R. J. (1991). The real world of teaching improvement. A faculty perspective. M THEALL & FRANKLIN (Eds.) **Effective Practices for Improving Teaching, New Directions for Teaching and Learning, Vol. 48**, (pp. 21-37). San Francisco: Jossey-Bass.

Miller, R. I. (1987). **Evaluation Faculty for Promotion and Tenure**. San Francisco: Jossey-Bass.

Miller, S. (1984). Student rating scales for tenure and promotion. **Improving College and University Teaching**, **32** (2), pp. 87-90

Millman, J. (1981). **Handbook of Teacher Evaluation**. Beverly Hills, CA: Sage.

Monroe, C. y Borzi, M. G. (1989). Methodological issues regarding student evaluation of teacher. A pilot study. **ACA Bulletin**, **70**, pp. 73-79.

Moses, I. (1986). Self and Student evaluation of academic staff. **Assessment and Evaluation in Higher Education**, pp. 76-78.

Muñiz, J. ; García, A. y Virgos, J. M. (1991). Escala de la Universidad de Oviedo para la evaluación del profesorado. **Psicothema**, **3** (2), pp. 269-281.

Murray, H. G.; Rushton, P. & Paunonen, S. V. (1990). Teacher Personality Traits and Student Instructional ratings in Six Types of University Courses. **Journal of Educational Psychology**, **82** (2), pp. 250-261.

Nimmer, J. G. y Stone, E. F. (1991). Effects of grading practices and time of rating on student ratings of faculty performance and student learning. **Research in Higher Education**, **32** (2), pp. 195-215.

Noser, T. C. ; Manakyan, H. y Tanner, J. R. (1996). Research productivity and perceived teaching effectiveness. A survey of economic faculty. **Research in Higher Education**, **37** (3), pp. 299-321.

O'connell, D. Q. y Dickenson, D. J. (1993). Student ratings of instruction as a function of testing conditions and perceptions amount learned. **Journal of Research and Development in Education**, **27** (1), pp. 18-23.

Overall, J. U. & Marsh, H. W. (1980). Students' s Evaluations of Instruction. A Longitudinal Study of Their Stability. **Journal of Educational Psychology**, **72**, pp. 321-325.

Palchik, N. S. (1988). Student assessment of teaching effectiveness in a multi-instructor course for multidisciplinary health professional student. **Evaluation and the Health Professions**, **11** (1), pp. 55-73.

Powell, R. W. (1977). Grades, learning, and student evaluation of instruction. **Research in Higher Education**, **7**, 193-205.

Prave, R. S. y Bavril, G. L. (1993). Instructor rating. controlling for bias from initial student interest. **Journal of Educational for Business**, **68** (2), pp. 362-366.

Prosser, M. & Trigwell, K. (1990). How will Future Academic be Evaluated? Using Student Study Strategies to **Check the Validity of Student Evaluations of Teaching Courses**. G. MULLIS (Ed)

Ramsden, P. (1991). A performance indicator of teaching quality in higher education. the course experience Questionnaire. **Studies in Higher Education, 16**, pp. 129-150.

Rindermann, H. & Schofield, N. (2001). Generalizability of Multidimensional Student Rating of University Instruction Across Courses and Teacher. **Research in Higher Education. Vol. 42**, 4, pp. 377-400.

Rodin, M. & Rodin, B (1972). Student Evaluation of Teachers. **Science, 177**, pp. 1164-1166.

Rutland, P. (1990). Some considerations regarding teaching evaluations. **Political Science Teacher, 3**, pp. 1-2.

Ryan, J. M. & Harrison, P. D. (1995). The Relationship Between Individual Instructional Characteristic and the Overall Assessment of Teaching Effectiveness Across Different Instructional Context. **Research In Higher Education, Vol. 36**, No 5, pp. 577-594

Ryans, D. G. (1960) **Characteristics of Teachers**. Washington, D. C.: American Council on Education.

Salvador, L. (1990). Los Docentes Universitarios Exitosos Desde La Perspectiva del Alumno. Su Caracterización Psicopedagógica. Tesis Doctoral. Universidad de Salamanca. España.

Saroyan, A. & Amundsen, Ch. (2001). Evaluating University teaching. Time to Take Stock. **Assessment & Evaluation in Higher Education, Vol 26**, (4), pp. 341-353

Seldin, P. (1984). Changing Practices in Faculty Evaluation. A Critical assessment and Recommendations for Improvement. San Francisco: Josser – Bass.

Seldin, P. (1993a). The use ad abuse of students ratings of professors. **The Chronicle of Higher Education**, p. 40

Seldin, P. (1993b). **Successful Use Of Teaching Portafolios**. Bolton, MA: Anker Publishing Co.

Shadish, W. (1998). **Some Evaluation Questions. Practical Assessment, Research & Evaluation, 6 (3)**. Recuperado el 14 de enero de 2001 de <http://ericae.net/pare/getvn.asp?v=6&n=3>].

Snyder, C. R. , & Clair, M. (1976). Effects of expected and obtained grades on teacher evaluation and attribution of performance. **Journal of Educational Psychology, 68**, 75-82.

Spencer, P. A. Y Flyr, M. L. (1992). **The Formal Evaluation As Ac Impetus To Classroom Change. Myths or Reality?**. Research/ Technical Report, Riverside, CA.

Tagomori, H. & Bishop, L. (1995). Student Evaluation of Teaching. Flaw Instruments. Thought and Action. **The National Education Association Higher Education Journal, 11**, pp. 63-78.

Tatro, C. N. (1995). Gender effect on student evaluations of faculty. **Journal of Research and Development in Education, 28 (3)**, pp. 169-173.

Tejedor, F. y Montero, L. (1990). Indicadores de la Calidad Docente para La Evaluación del Profesor Universitario. **Revista Española De Pedagogía, año XLVIII, N° 186**, mayo-agosto, pp. 259-279.

Timpson, W. W. Y Andrew, D. (1997). Rethinking student evaluation and the improvement of teaching. instrument for change at the University of Queensland. **Studies in Higher Education, 22 (1)**, pp. 55-65.

Ting, K. F. (2001). A Multilevel Perspective On Student Ratings of Instruction. Lessons From the Chinese Experience. **Research in Higher Education. Vol. 41, 5**, pp. 637-653.

Trigwell, K. & Prosser, M. (1996). Changing Approaches to Teaching. A Relational Perspective. **Studies in Higher Education, 21**, pp. 275-284.

Vandewalle (1997). Development and validation of a work domain goal orientation instrument. **Educational and Psychological Measurement, 57 (6)**, pp. 995-1015.

Vasta, R. , & Sarmiento, R. F. (1979). Liberal grading improves evaluations but not performance. **Journal of Educational Psychology, 71**, 207-211.

Villa, A. Y Morales, P. (1993). **La Evaluación Del Profesor. Una Visión De Los Principales Problemas Y Enfoques De Diversos Contextos.** Vitoria: Departamento de Educación, Universidades e Investigación. Gobierno Vasco.

Vu, T. R. ; Marrito, D. J. ; Stratos, G. A. Y Litzelman, D. K. (1997). Prioritizing areas for faculty development of clinical teachers by using student evaluations for evidence- based decisions. **Academic Medicine**,**72** (10), pp. 57-59.

Wachtel, H. K. (1998). Student Evaluation of College Teaching Effectiveness. A Brief Review. **Assessment & Evaluation In Higher Education**, **23** (June), 191-211.

Weinbach, R. W. (1988). Manipulation of student evaluations. No laughing Mater. *Journal of Social Work Education*, 24 (1), pp. 37-34.

Wigington, H. ; Tollefson, N. Y Rodríguez, E. (1989). Student's ratings of instructor visited. Interactions among class and instructor variables. **Research in Higher Education**, **30** (3), pp. 331-334.

Wilkerson, D. ; Rogers, M. A. Y Maughan, R. (2000). Validation of Student, Principal, and Self Ratings in 360 Feedback for Teacher Evaluation. **Journal of Personnel Evaluation in Education**, **14** (2), pp. 179-192

Wilson, K. L.; Lizzio, A. & Ramsden, P. (1997). The Development , Validation, and Application of the Course Experience Questionnaire. **Studies in Higher Education**, **22** (1) , pp. 33-52.

Worthington, A. G. , & Wong, P. T. P. (1979). Effects of earned and assigned grades on student evaluations of an instructor. **Journal of Educational Psychology**, **71**, 764-775.

Worthington, A. C. (2002). The impact of student perception and characteristics on teaching evaluation. a case study in finance education. **Assessment & Evaluation in Higher Education**, **27** (1), pp. 49-64.

Zoller, U. (1992). Faculty Teaching Performance Evaluation in Higher Science Education. Issues and Implications. **Science Education**, **76** (6), pp. 673-684.