

ANÁLISIS DIMENSIONAL DE LAS OPINIONES DE LOS ALUMNOS UNIVERSITARIOS SOBRE SUS PROFESORES: COMPARACIÓN ENTRE TÉCNICAS PARAMÉTRICAS Y NO-PARAMÉTRICAS

Apodaca, P.

Universidad del País Vasco

Grad, H.

Universidad Autónoma de Madrid

RESUMEN

Se discuten las posibilidades y limitaciones de las funcionalidades sumativa y formativa en el uso de las encuestas de opinión del alumnado sobre la calidad docente de su profesorado. Se relacionan esas funcionalidades con el carácter unidimensional o multidimensional de tales encuestas. La revisión de los estudios estructurales realizados por diversos autores muestra una gran concordancia entre las dimensiones en ese contexto. Se discuten las ventajas y limitaciones de técnicas estadísticas paramétricas y no paramétricas (LISREL vs. SSA) para analizar la estructura o dimensionalidad del constructo de competencia docente. Estas técnicas son aplicadas al análisis de datos de dos amplias muestras de dos universidades, con dos instrumentos distintos. Los resultados encontrados mediante LISREL concuerdan con estudios previos similares. Los resultados encontrados mediante SSA aportan evidencia para una reconceptualización no lineal del constructo de competencia docente que integra adecuadamente las tradicionales conceptualizaciones unidimensionales y multidimensionales.

Palabras clave: Educación Superior, opinión de los alumnos sobre sus profesores, dimensionalidad, análisis factorial confirmatorio, análisis estructural no paramétrico.

* Instituto Ciencias Educación. Universidad del País Vasco. 48940 Leioa (Bizkaia). c.e.: plpapurp@lg.ehu.es

** Departamento de Antropología Social. Facultad de Filosofía y Letras. Universidad Autónoma de Madrid. Ciudad Universitaria de Cantoblanco. 28049 Madrid. c.e.: hector.grad@uam.es

ABSTRACT

The possibilities and limitations of the summative and formative functionalities in the use of the students ratings are discussed. These functionalities are closely related to the unidimensional or multidimensional character of such surveys. A revision of the dimensions proposed by diverse authors were made and a great agreement found. The contributions and limitations of parametric and nonparametric statistical techniques (LISREL versus SSA) are discussed in order to analyze the structure and dimensionality of the construct of educational competence. These techniques were applied to ample samples of two universities answering to two different instruments. Results found by means of LISREL are consistent with previous similar studies. Results found by means of SSA bring out evidence for a nonlinear reconceptualization of the construct educational competence that suitably integrates the traditional unidimensional and multidimensional conceptualizations.

Key words: Higher Education, student's ratings of teaching, dimensionality, lisrel, smallest space analysis.

I. INTRODUCCIÓN

Las encuestas de opinión del alumnado sobre la docencia del profesorado universitario son de uso general en la mayoría de las universidades del país. Los momentos iniciales de esta práctica en los años ochenta correspondían al movimiento general por la renovación y democratización de la universidad española acompañado de una fuerte presión del movimiento estudiantil. Estos movimientos llegaron a conseguir en la mayoría de los casos que los Estatutos, reglamentos o normas recogieran explícitamente la obligatoriedad de realizar este tipo de consulta (Mateo, 1987; Tejedor, 1993). De esta manera, a comienzos de los noventa la gran mayoría de las universidades tenían formalizadas encuestas de opinión de los alumnos sobre sus profesores. Debe constatar, sin embargo, una gran diversidad en el compromiso e implicación de la comunidad universitaria con esta evaluación. Esta diversidad se reflejó en los mecanismos adoptados para la evaluación —desde la opcionalidad abierta a la voluntariedad del profesorado, pasando por la pseudo-obligatoriedad sin preocupación real por garantizar la cobertura general de la encuesta, hasta la obligatoriedad establecida en estatutos y asegurada por los procedimientos de aplicación—.

No obstante, paralelamente a la generalización de estas prácticas, se produce un progresivo desinterés, en ciertos casos desengaño, por parte de los colectivos estudiantiles. Ello podría estar asociado a varios factores entre los que cabe destacar los siguientes.

En primer lugar, la ambivalencia institucional con respecto al valor y la utilidad evaluativa desde el punto de vista de los alumnos (como consecuencia, por un lado, de la actitud favorable a la evaluación para la mejora y, por otro lado de la dificultad y/o resistencia para implantar los mecanismos consiguientes de información, toma de decisiones y promoción de la mejora docente). En segundo lugar, en parte por el factor anterior y en parte por la dilación del mecanismo de información y toma de decisio-

nes, surge cierta dificultad de visualizar el impacto de estas prácticas evaluativas sobre la calidad docente, tanto para el profesorado como para el alumnado, que repercute en la desmotivación de ambos colectivos hacia la realización de la encuesta. Finalmente, cabe citar la deriva de los proyectos de cambio hacia proyectos de gestión de la universidad existente tras la LRU, correlativa con el desgaste del profesorado y la desmovilización del alumnado, y la disminución del compromiso con el cambio por parte de estudiantes e institución.

En cualquier caso, para discutir apropiadamente la dimensión y trascendencia de este mecanismo de evaluación, es necesario señalar que la opinión del alumno es, necesariamente, una visión parcial e incompleta del constructo de 'competencia docente'. Por su propia posición en el proceso de enseñanza-aprendizaje, los estudiantes no serán una fuente válida y fiable de información sobre aquellos aspectos de la enseñanza que no pueden observar sistemáticamente o sobre aquellos aspectos donde el conflicto de intereses o motivaciones pueda sesgar claramente sus percepciones y valoraciones. Por lo tanto, el uso de las mencionadas encuestas debe ser necesariamente integrado en una visión más amplia de la evaluación del profesorado que contemple otros criterios, fuentes e instrumentos (De Miguel, 1998; De Miguel et al., 1991).

Asimismo, por sus implicaciones metodológicas, debe tenerse en cuenta que la estabilización de estos procesos se ha asociado progresivamente a su uso sumativo en detrimento de la funcionalidad eminentemente formativa para la que originalmente se diseñaron los instrumentos y procedimientos de aplicación de las encuestas y de información a las partes interesadas (órganos de gobierno, profesorado, estudiantes). De esta manera, queda pendiente el análisis de la adecuación de instrumentos y procedimientos a las nuevas funcionalidades sumativas de estas encuestas.

La Ley Orgánica de Universidades (BOE 24/12/01) profundiza en la orientación sumativa de la evaluación del profesorado al establecer procedimientos evaluativos para el acceso a las nuevas figuras contractuales de profesorado y, sobre todo, establece la posibilidad de otorgar complementos salariales en función de la calidad o productividad.

Paralelamente a esta consolidación de la vertiente sumativa en la evaluación del profesorado, numerosas universidades han puesto en marcha importantes proyectos de formación inicial y continua para la capacitación docente de su personal. Por lo tanto, se consolida también la necesidad de desarrollar el uso formativo de las estrategias evaluativas desde una orientación de desarrollo profesional del profesorado universitario.

En esta situación, es previsible que las encuestas de opinión del alumnado sobre la calidad docente de su profesorado tengan un papel creciente en los procesos de formación, selección, promoción e incentivación salarial. Son evidentes las amenazas y oportunidades que ello puede representar. El correcto uso de dichas encuestas requiere prestar especial cuidado tanto a los aspectos sustantivos de la evaluación del personal como a la calidad de los instrumentos utilizados y a los procedimientos de explotación de los resultados e información.

Otra cuestión a tener en cuenta es el carácter unidimensional o multidimensional de los cuestionarios utilizados en estas encuestas o del constructo mismo de 'calidad

docente desde la perspectiva del alumno'. Esta cuestión tiene implicaciones directas para las funcionalidades sumativa y formativa de la evaluación, pues plantea el dilema de tener que elegir entre una puntuación general única de 'competencia docente' o varias puntuaciones de las distintas dimensiones.

La funcionalidad formativa de la evaluación sugiere la conveniencia de un enfoque multidimensional para evaluar cada uno de los distintos aspectos de la actuación docente con la precisión necesaria para definir pautas de intervención ajustadas a las necesidades individuales. En este contexto, una única medida general de la actuación carecería de utilidad ya que no haría explícito el 'perfil de calidad' y, por lo tanto, los aspectos de la actuación docente merecedores de mejora.

Por el contrario, el enfoque unidimensional parece tener ventajas frente al multidimensional para la funcionalidad sumativa de la evaluación. Autores como Abrami y Apollonia (1990) defienden con claridad el uso de una única medida general bien sea a partir de ítems de valoración general o a partir de un cuidadoso y selectivo promedio ponderado de ítems/dimensiones sobre conductas/competencias específicas. En un trabajo posterior son aún más restrictivos recomendando que para objetivos sumativos se utilicen únicamente ítems de valoración general de la efectividad del profesor (Apollonia y Abrami, 1997). Desde la validez de contenido, determinadas dimensiones de la actuación docente efectiva se verían afectadas por variables ajenas a la responsabilidad del profesor. Como ejemplo puede aportarse las limitaciones en la dimensión de 'interacción con los alumnos' en clases grandes. También se puede hacer referencia aquí a la escasa relación de las opiniones de los alumnos sobre determinadas dimensiones de la actuación docente con los resultados académicos (usados como medida del aprendizaje efectivo del alumno), en contraposición a las correlaciones importantes de los ítems de valoración general con el citado aprendizaje (Cohen, 1981). Por último, el uso de criterios multidimensionales es más complejo y arriesgado para gestores inexpertos en enseñanza y evaluación, pues éstos pueden ponderar e integrar estos criterios de forma inadecuada y establecer juicios comparativos descontextualizados.

Obviando estas cuestiones, el debate sobre el carácter unidimensional o multidimensional de estos cuestionarios ha estado muy ligado, quizás inapropiadamente, a los resultados empíricos obtenidos mediante análisis factorial exploratorio y, más recientemente, confirmatorio. Sin embargo, son bien conocidas las limitaciones de estas y otras técnicas para generar teoría o identificar aspectos sustantivos de un constructo (Elejabarrieta y Perera, 1989). Asimismo, el carácter 'confirmatorio' de estos análisis empíricos se ha demostrado dudoso dada la gran flexibilidad que aporta permitiendo e incluso empujando a una estrategia analítica de aproximaciones sucesivas al modelo a contrastar finalmente. La propia técnica estadística (o la necesaria especificación de los parámetros del modelo) limita las soluciones posibles y es muy restrictiva en sus condiciones de aplicación. Finalmente, los coeficientes de ajuste manejados o las pruebas de contraste de hipótesis son muy vulnerables a múltiples factores (Apodaca y Páez, 1992). En síntesis, ninguna técnica de análisis empírico puede sustituir la necesaria elaboración teórica previa del constructo de actuación docente.

En este ámbito, las técnicas no-paramétricas de escalamiento multidimensional han sido escasamente utilizadas. La ausencia de condiciones previas con respecto a las distribuciones de las variables, su flexibilidad a la hora de representar estructuras latentes lineales (dimensionales) y no lineales (como el orden circular) y su uso combinado con la Teoría de las Facetas (Borg & Shye, 1995; Canter, 1985; Dancer, 1990; Guttman & Greenbaum, 1998; Levy, 1994) para formalizar sistemáticamente las hipótesis induce a pensar que estas técnicas pueden aportar claridad conceptual y viabilidad operativa en el uso de estos cuestionarios para fines tanto sumativos como formativos.

El presente trabajo centra sus esfuerzos precisamente en esta cuestión. Se trata de comparar las soluciones que ofrece el análisis factorial confirmatorio lineal (modelo de medida) con el análisis estructural no-paramétrico. Esta comparativa puede aportar elementos nuevos y significativos al debate sobre el carácter dimensional de estos cuestionarios y, por lo tanto, a la especificación de procedimientos para un uso más adecuado de sus resultados.

En el ámbito internacional existen numerosos trabajos sobre este tema tanto desde técnicas paramétricas como no-paramétricas aunque las primeras son francamente dominantes. Sin embargo, son muy escasos los realizados en el ámbito de las universidades españolas y casi en su totalidad utilizan el análisis factorial exploratorio y confirmatorio. Cabe destacar los estudios de F.J. Tejedor (1993) y de este autor y colaboradores (1987, 1988) como los de mayor impacto en este ámbito. Son destacables también otros trabajos en colaboración con autores anglosajones como los de Tourón (1989). Asimismo, trabajos más recientes, como los de García Ramos (1997), introducen el análisis factorial confirmatorio actualizando en lo metodológico estas investigaciones. El trabajo de Villa y Morales (1993) resulta también una buena revisión de diversas estructuras dimensionales en una serie de instrumentos manejados en la evaluación del profesorado.

En primer lugar se presentará con brevedad una revisión de las dimensiones que suelen aparecer en estos instrumentos. Podrá observarse una cierta estabilidad de los resultados (que puede deberse, en parte, a la similitud del contenido de los cuestionarios más difundidos) así como una cierta dependencia de los mismos hacia la técnica estadística empleada o hacia los parámetros especificados. A continuación, se presentará un análisis de las características, posibilidades y limitaciones de las técnicas estadísticas paramétricas y no-paramétricas utilizadas. Finalmente, se presentan y comentan los resultados encontrados con ambas técnicas aplicadas a dos cuestionarios de 2 universidades diferentes.

2. DIMENSIONALIDAD DE LOS INSTRUMENTOS

Una visión esquemática de esta cuestión podría establecer como polos opuestos la visión unidimensional frente a la multidimensional. El hecho de que ambas posturas puedan aportar resultados empíricos que las avalen complica el debate. Las razones para este hecho pueden ser de diverso tipo. Desde la propia técnica estadística utilizada, pasando por el contenido concreto de cada cuestionario, para finalizar en la mayor o menor longitud de los instrumentos.

Una rápida revisión de algunos de los cuestionarios aplicados en las universidades españolas hace evidente una gran similitud entre ellos. Su longitud tiende a establecerse por debajo de los 20 ítems y el contenido de los mismos es equivalente si no idéntico en su gran mayoría. Las razones pueden ser varias pero el factor clave es que se nutren de fuentes comunes entre las que cabe destacar los estudios pioneros de Aparicio, Tejedor y SanMartín (1982) continuados por Tejedor (op. cts.) y la posterior colaboración e intercambio de experiencias entre las unidades de evaluación.

La similitud entre los instrumentos permite esperar una gran semejanza en su estructura dimensional. Uno de los objetivos colaterales del presente trabajo es verificar esa semejanza estructural comparando los cuestionarios aplicados en dos universidades del Estado.

Las revisiones y estudios del ámbito anglosajón aportan resultados contradictorios. Por un lado, tenemos los trabajos que abogan con claridad por un enfoque multidimensional. Destacan entre estos trabajos por su amplitud y continuidad en el tiempo los trabajos de Marsh (Marsh y Hocevar, 1984; Marsh, 1988; Marsh y Hocevar, 1991; Marsh y Dunkin, 1992; Marsh y Roche, 1997, entre otros). Defiende el amplio acuerdo entre numerosos investigadores en el carácter complejo y multidimensional de la actividad docente y defiende especialmente que las evaluaciones de tipo formativo-diagnóstico debieran reflejar esta multidimensionalidad. Su propuesta dimensional basada en los trabajos arriba referidos arroja una estructura de 9 factores principales que pueden integrarse en cuatro factores de segundo orden.

TABLA 1
ESTRUCTURA DIMENSIONAL DEL 'STUDENT'S EVALUATION
OF EDUCATIONAL QUALITY' (SEEQ) DE H. W. MARSH

Factores de 1er. orden	Factores de 2º orden
Aprendizaje	I. Presentación
Entusiasmo del profesor	
Organización/Claridad	
Programa	
Interacción con el grupo	II. Rapport
Interacción individual	
Exámenes/Calificaciones	III. Materiales del curso
Trabajo extraclases/ Lecturas	III / IV
Carga de trabajo / Dificultad	IV. Carga de trabajo

Esta estructura, en su versión de 9 factores o de cuatro factores de segundo orden, constituye una referencia ineludible en la construcción y uso de estos instrumentos. Sin embargo, podría criticarse a estos trabajos su excesiva dependencia de la técnica estadística en detrimento del análisis lógico y conceptual. En definitiva, sería cuestionable el uso del análisis factorial como evidencia fundamental para apoyar la validez

de constructo de estos cuestionarios. Esta validez de constructo debiera apoyarse más en la teoría y contar con investigación empírica adicional al análisis factorial.

Otros trabajos, entre los que cabe destacar a Feldman (1989, 1996) aportan un enfoque menos empírico y más conceptual, identificando de esta manera 20 dimensiones básicas que podrían agruparse —también mediante análisis lógico— en tres grandes dimensiones: presentación, regulación, facilitación.

En estos y otros trabajos puede observarse un alto nivel de covariación entre el conjunto de ítems y entre los factores de primer orden extraídos. En determinada medida se aprecia con claridad la posibilidad de plantear un factor general. En muchos casos, como planteamiento alternativo se establecen estructuras multifactoriales de segundo orden que bien pudieran integrarse en un único factor general. De esta manera, la alternativa metodológica de considerar como unidimensionales estos cuestionarios puede tener un apoyo empírico además del conceptual o lógico.

Los autores que quizás con mayor claridad han señalado la fuerte unidimensionalidad de estos cuestionarios han sido Abrami y Apollonia (Abrami, 1989; Abrami y Apollonia, 1990; Apollonia y Abrami, 1997; Abrami y Apollonia, 1999). Por un lado han distinguido con claridad los usos formativos y sumativos de la evaluación a la hora de contemplar diferentes dimensiones o una única valoración general. Por otro lado, señalan que aunque la docencia es multidimensional, los ítems que componen estos cuestionarios están saturados por un factor global bastante potente. Desde la vertiente más empírica estos autores reanalizan estudios factoriales realizados por otros autores empleando diversos instrumentos. De los siete cuestionarios analizados, en seis de ellos encuentran —mediante análisis de componentes principales y sin rotación— un primer factor general que explica más del 30% de la varianza siendo el resto de los factores, en la mayoría de los casos comparativamente residuales. El uso de una rotación posterior podría dar lugar a diversos factores de potencia explicativa más equilibrada pero aún siendo soluciones matemáticamente equivalentes cabe plantear si la solución rotada no constituye un 'artefacto' de la técnica estadística más que algo sustantivo del propio constructo. Así, niegan que los resultados empíricos apoyen la multidimensionalidad de estos instrumentos, más bien son evidencia de la existencia de un componente o dimensión general.

Cabe señalar aquí la similitud de este debate con el producido en torno a las teorías de la inteligencia a mediados del siglo pasado. El uso de diferentes métodos de extracción, del tipo o ausencia de rotación, del conjunto de parámetros especificados en el modelo, dan lugar a diferentes soluciones con los mismos datos. De alguna manera, se evidencian las limitaciones de las técnicas estadísticas como elemento principal en la construcción de teoría. El uso del análisis factorial confirmatorio tampoco ha contribuido a la resolución de este problema.

Finalmente, cabe plantearse un postura práctica y operativa. Conceptualizar el constructo como unidimensional y, a la vez, con dimensiones específicas de menor peso. Esta postura está implícita en los trabajos tanto de Marsh (op. cts.) como de Abrami y Apollonia (op. cts.), y es defendida explícitamente por McKeachie (1997). El uso para fines sumativos o formativos de estos cuestionarios daría lugar a diferentes

TABLA 2
 ESTRUCTURA DIMENSIONAL (APODACA Y RODRÍGUEZ, 1999):
 ÍTEMS REPRESENTATIVOS

Dimensión 1: *Cumplimiento formal del profesor/a*

- Asiste regularmente a clase
- Cumple debidamente su horario
- En las horas de tutoría está disponible en su despacho

Dimensión 2: *Desarrollo del programa de la asignatura*

- Comenta el programa de la asignatura al inicio del curso
- Considero que desarrollará todo el programa

Dimensión 3: *Dominio, organización y claridad*

- Parece dominar la asignatura que imparte
- Sus clases están bien preparadas
- Responde con exactitud y precisión a las preguntas que se le hacen
- Explica con claridad
- Consigue que el alumnado se interese por la asignatura
- Se preocupa de que sus clases sean buenas
- Se preocupa por renovar los contenidos y los métodos de enseñanza

Dimensión 4: *Interacción profesor-alumno*

- Intenta que el alumnado participe en las clases
- Fomenta la crítica de sus propias ideas
- Dialoga con el alumnado sobre la marcha de la clase
- Tiene en cuenta la opinión del alumnado a la hora de decidir las cuestiones relacionadas con la marcha de la asignatura

Dimensión 5: *Recursos metodológicos-didácticos*

- Apoya sus explicaciones con esquemas, gráficos, medios audiovisuales, etc.
- Los materiales utilizados y/o recomendados (apuntes, libros,...) son de gran ayuda
- Los problemas, ejemplos o prácticas que plantea están bien pensados en el contexto de la asignatura

Dimensión 6: *Exámenes y calificaciones*

- Sus exámenes se ajustan a los contenidos impartidos
- Las calificaciones obtenidas por alumnado se ajustan a sus conocimientos
- Se muestra dispuesto a comentar y razonar la nota obtenida por los/as alumnos/as que reclaman revisión de examen

Dimensión 7: *Valoración global*

- Teniendo en cuenta las limitaciones, pienso que el/la profesor/a que imparte esta asignatura es un buen/a profesor/a
- Me siento satisfecho/a asistiendo a sus clases
- He aprendido mucho cursando esta asignatura

procedimientos de explotación con el objetivo de obtener una única puntuación general o diversas puntuaciones más específicas.

En un estudio previo (Apodaca, 1995), mediante análisis factorial confirmatorio (LISREL), se intentó contrastar el mayor o menor ajuste a los datos de las soluciones unidimensionales y multidimensionales. La solución unidimensional presentaba índices de ajuste considerablemente menores. Sin embargo, este hecho está fuertemente asociado a su simplicidad ya que con esta técnica las soluciones más parsimoniosas son las menos ajustadas, siendo evidente la existencia de estructuras de covariación independientes para subconjuntos de ítems. La solución multidimensional presentaba índices de ajuste muy aceptables y ventajas operativas evidentes en la elaboración de informes de resultados y en su uso formativo.

En un trabajo posterior (Apodaca y Rodríguez, 1999) se replicó este mismo estudio encontrando resultados similares a los del estudio anterior y a los reflejados en las revisiones internacionales y nacionales. A continuación se presenta la estructura dimensional que podríamos calificar de 'tipo' con la especificación de algunos de los ítems más característicos de dichas dimensiones.

Es necesario señalar que la primera dimensión —cumplimiento formal— no debiera considerarse propiamente como inherente a la calidad o competencia docente del profesor. Se trata de un conjunto de ítems que tratan de verificar los requisitos puramente formales y administrativos como son la asistencia a clase y el cumplimiento del horario. Su habitual inclusión en este tipo de cuestionarios obedece más bien a la ausencia de mecanismos administrativos ad-hoc para realizar esta verificación. Por lo tanto, podemos considerar esta dimensión como ajena al propio constructo de competencia docente del profesor.

En cuanto a la última dimensión —valoración global— se trata de ítems donde se recoge una impresión y opinión general sobre el profesor. Por su carácter integrador, la inclusión de estos ítems en análisis dimensionales es parcialmente redundante con los aspectos específicos y, por lo tanto, robustecen las soluciones unidimensionales.

Si dejamos al margen estas dos dimensiones, se hace evidente la similitud o al menos paralelismo entre esta estructura dimensional y la defendida en los estudios más reconocidos de ámbito internacional. La comparación con la Tabla 1 permite observar la concordancia entre las dimensiones encontradas y las dimensiones de Marsh (op. cit.): la segunda y la sexta de las dimensiones encontradas ('Desarrollo del programa' y 'Exámenes y calificaciones') se solapan, respectivamente, con la cuarta y la séptima dimensiones de Marsh ('Programa' y 'Exámenes/calificaciones'). La dimensión etiquetada 'Dominio, organización y claridad' incluye aspectos de las dimensiones 'Organización/Claridad' y 'Entusiasmo del profesor' de Marsh. Las dimensiones de 'Interacción individual' e 'Interacción grupal' se presentan unidas en nuestros instrumentos en una única dimensión de 'Interacción profesor-alumno'. Aunque la dimensión etiquetada 'Recursos metodológicos-didácticos' pueda corresponderse parcialmente con la dimensión 'Trabajo extraclases/Lecturas', tanto esta dimensión como la de 'Carga de trabajo/Dificultad' planteadas por Marsh son de difícil homologabilidad en el contexto de enseñanza-aprendizaje de nuestro sistema universitario. Por

último, ningún ítem del cuestionario analizado refleja claramente la primera dimensión de Marsh ('Aprendizaje').

3. ANÁLISIS ESTRUCTURAL PARAMÉTRICO Y NO-PARAMÉTRICO

3.1. Análisis dimensional paramétrico

En los últimos diez años se ha producido un enorme auge en el uso de modelos de ecuaciones estructurales (Batista y Coenders, 2000) una de cuyas herramientas más extendidas es el LISREL. Otras herramientas equivalentes como el EQS y, más recientemente, el AMOS (como nombre comercial del LISREL una vez que SPSS se hizo con la licencia) compiten en el mercado ofreciendo algunas variantes. La progresiva simplificación en el manejo de estos programas ha permitido un uso cada vez más extendido. Dentro de esta técnica estadística puede distinguirse el modelo estructural y el modelo de medida. En este estudio trabajaremos únicamente con el modelo de medida.

Esta técnica se enmarca plenamente dentro de la familia de las técnicas paramétricas por los supuestos que realiza y por la posibilidad de realizar estimaciones del ajuste y contrastes de hipótesis de acuerdo a distribuciones de probabilidad conocidas como el Chi Cuadrado.

El uso de esta técnica, que podemos denominar LISREL para simplificar, presenta algunas limitaciones importantes que deben considerarse al usarse como análisis factorial confirmatorio. En general, puede observarse con cierta frecuencia una escasa atención a los supuestos básicos de la técnica y a los límites en la interpretación de los resultados.

Entre las asunciones básicas cabe recordar algunas de ellas. Los modelos han de ser completos en el sentido de contener una muestra representativa y suficiente del universo de ítems del constructo a medir. La inclusión o exclusión de elementos característicos puede alterar la estimación de los parámetros del conjunto. La linealidad de las relaciones entre las variables observadas y de éstas con las latentes es otro de los supuestos de la técnica aunque puede considerarse una buena aproximación a la mayoría de las funciones (Satorra y Stronkhorst, 1984). Finalmente, la normalidad de las distribuciones es otra de las asunciones.

Sin embargo, para el manejo e interpretación de los resultados, en especial de la prueba de hipótesis de ajuste del modelo mediante Chi Cuadrado, son de especial trascendencia otras propiedades de esta técnica. En resumen, cabe señalar que un modelo tenderá a ser estadísticamente significativo (no ajustado a los datos) cuanto mayor sea el tamaño de la muestra (Maruyama, 1998), cuanto más restrictivo-parsimonioso sea el modelo y cuanto mayores sean las covarianzas entre las variables. En definitiva, con esta técnica se favorecen claramente los estudios que manejen pequeñas muestras, postulen modelos complejos y se basen en correlaciones débiles (Apodaca y Paéz, 1992). Esta situación aparentemente paradójica requiere un uso cuidadoso de la técnica y, en especial, de las pruebas de hipótesis del ajuste del modelo.

También deben tenerse en cuenta otras cuestiones de tipo más metodológico que estadístico. Por un lado, es preciso recordar que esta técnica, como cualquier prueba de hipótesis, no permite confirmar un modelo determinado sino únicamente dejar de rechazarlo, es decir, su compatibilidad con los datos y su pertenencia a la familia de modelos alternativos igualmente compatibles con los datos. La conocida falacia 'nominalista' es otra cuestión que conviene recordar. Es decir, conseguir poner nombre a un factor o variable latente no significa que tenga un carácter real o sustantivo. La técnica puede garantizar que un conglomerado de ítems covarían pero no que dicha covariación sea efecto de un factor o variable latente con existencia real.

Ahora bien, para el trabajo que aquí nos ocupa existen dos aspectos relativos a esta técnica estadística que son a la vez su fortaleza y su debilidad/limitación. Se trata por un lado de la manipulabilidad del ajuste del modelo de medida planteado y, por otro lado, de la estructura jerárquica impuesta a dicho modelo.

Respecto a la manipulabilidad del modelo, es preciso señalar primeramente y con contundencia que la técnica LISREL es sólo virtualmente confirmatoria. La práctica real la asocia con claridad a una técnica exploratoria. Esta técnica plantea infinidad de posibilidades en la definición de los parámetros del modelo y la teoría difícilmente puede establecer con claridad una parte importante de estas definiciones; bien sea por el escaso desarrollo de la teoría, bien por la irrelevancia o marginalidad de determinados aspectos. De esta manera, en la práctica real, los modelos no se contrastan una sola vez sino que van modificándose en sucesivas pruebas. En estas 'modificaciones' de los parámetros juegan un papel fundamental los coeficientes de ajuste logrados. En definitiva, se va alterando el modelo, a veces en lo sustantivo, en función del ajuste logrado (Batista y Coenders, 2000). De alguna manera son los datos los que construyen el modelo aunque la teoría haya jugado un papel importante en la definición del modelo inicial. El grado en el cual se es respetuoso con la teoría en las sucesivas modificaciones del modelo queda en manos de la profesionalidad del investigador. Todo ello nos sitúa con claridad en una estrategia más bien exploratoria en lugar de confirmatoria. El carácter confirmatorio no lo dará tanto la propia técnica como la metodología usada, por ejemplo acudir a estrategias tales como la partición de la muestra empleado, la verificación del modelo con otras muestras-situaciones-investigadores, etc. (Maruyama, 1998).

El aspecto que mayor flexibilidad-manipulabilidad presta a la técnica es la definición de los parámetros de los errores de las variables observables (ítems). Estos errores normalmente pueden definirse como independientes entre sí. Sin embargo, cabe plantearse fuentes de error o sesgo que afectan a más de un ítem produciendo correlación entre los errores. Asimismo, cabe plantearse la presencia de factores o variables latentes que afectan a determinados ítems pero que no forman parte de lo sustantivo del constructo. En definitiva, puede ser razonable o justificado el postular determinar estructuras de covariación entre los errores (Byrne, 1998).

Sin embargo, definiendo los errores como intercorrelacionados en función de los índices de ajuste en modelos previos, el modelo se ajustará progresivamente y la varianza no explicada por los factores o variables latentes quedará recogida en dichas intercorrelaciones dándose la paradoja de que el modelo podría llegar a ajustarse más

por dichas correlaciones entre los errores que por los factores o variables latentes postulados.

La segunda cuestión que le presta especial interés a la técnica a la vez que la limita considerablemente en su cobertura es que sólo puede manejar estructuras jerárquicas. En definitiva, sólo sería apropiada para constructos cuya estructura fuera de 'árbol'. Así, al igual que el análisis factorial exploratorio clásico, las variables observables (ítems) se agrupan en factores de primer orden y éstos en factores de segundo orden y así sucesivamente. Ahora bien, esa conceptualización lineal de los constructos es sólo una de las posibilidades, ya que se podrían postular estructuras de otro tipo; como polares, circulares o en anillo, etc.

3.2. Análisis estructural no-paramétrico

Otras técnicas estadísticas tienen un carácter menos restrictivo y plantean nuevas posibilidades para la conceptualización y uso de estas encuestas. Entre ellas podemos destacar el '*Smallest Space Analysis*' (SSA). El SSA es una técnica ordinal de escalamiento multidimensional desarrollada por Guttman (1968; Lingoes & Guttman, 1967 - Levy, 1994, ha compilado una selección de las principales contribuciones de Guttman a la metodología de la investigación social). Esta técnica representa espacialmente la matriz de coeficientes de asociación entre las variables, de forma que las distancias entre los puntos sean una función monótona de esos coeficientes y la matriz de distancias obtenida tenga la menor dimensionalidad teóricamente justificada (para una introducción general al análisis multidimensional, ver Davison (1983), Dillon & Goldstein (1984), para una introducción al SSA, ver Canter, (1985), Borg & Lingoes (1987) y Shye & Elizur (1994).

Es decir, las variables son representadas como puntos en un espacio multidimensional en el que las distancias entre los puntos son una función inversa de sus relaciones empíricas (medidas por cualquier coeficiente de asociación o de distancia). Cuanto mayor es la asociación empírica entre dos variables, más próximas serán sus ubicaciones en la representación (Canter; 1985; Davison, 1985; Dillon & Goldstein, 1984; Shye, 1988; comparan SSA con otros métodos de análisis multidimensional, de factores y de conglomerados).

Siguiendo las propuestas de Guttman, el SSA se utiliza, generalmente, en combinación con la Teoría de las Facetas (ver Borg, 1981; Borg & Groenen, 1997; Borg & Lingoes, 1987; Borg & Shye, 1995; Canter, 1985; Dancer, 1990; Shye & Elizur, 1994). La Teoría de las Facetas es un instrumento lógico para formalizar sistemáticamente los aspectos empíricos de la teoría, facilitando una vinculación estrecha entre el marco teórico, el diseño de los instrumentos y el análisis de los datos. Esa estrecha relación resulta muy fructífera para la investigación, en tanto obliga al investigador a hacer explícitos a priori los aspectos y la dimensionalidad del constructo-disminuyendo el riesgo de interpretar ad-hoc resultados circunstanciales (Borg & Shye, 1995; Dancer, 1990; Guttman & Greenbaum, 1998).

Conceptualmente, el SSA supone la existencia de un universo continuo de contenido, del que las variables estudiadas representan una muestra concreta de ítems de ese

universo. Esta muestra se genera, en el marco de la Teoría de las Facetas, aplicando una frase de definición —*mapping sentence*— que especifica los elementos de las facetas de contenido que serán estudiadas.

Se espera que las variables teóricamente relacionadas converjan empíricamente permitiendo la partición del espacio del SSA en regiones disyuntas (en este contexto, una región es definida como cualquier subconjunto de variables representando un mismo elemento de alguna faceta de la frase de definición) y que las relaciones estructurales se reflejen en la configuración espacial de esas regiones. Es decir, las relaciones estructurales esperadas entre los elementos de las facetas son traducidas a hipótesis sobre la configuración de su representación espacial.

Por lo tanto, son esas configuraciones de puntos y de regiones en el espacio las que se someten a interpretación para verificar la teoría (un enfoque de «verificación de la configuración» en términos de Davison, 1983). Sólo esas configuraciones (las posiciones relativas de los puntos) tienen significado sustancial y pueden ser interpretadas, pues los ejes del SSA son arbitrarios. El SSA proporciona una representación gráfica de los coeficientes originales como función de las distancias en la representación (*diagrama de Shepard*) y un 'Coeficiente de Alienación' que refleja el grado de ajuste de la función (entre 0 y 1; 0 indica un ajuste perfecto y un valor inferior a ,20 es considerado adecuado). Sin embargo, este coeficiente no implica la validez estadística del resultado, que sólo es asegurada por la reiteración de evidencias. La significación del resultado depende fundamentalmente de la posibilidad de interpretar sustancialmente la representación espacial aplicando la teoría previa.

La Teoría de las Facetas permite también sistematizar la traducción de las hipótesis estructurales en hipótesis regionales y su posterior interpretación. Se espera que las relaciones entre las facetas y sus componentes se reflejen en diferentes estructuras de la matriz de coeficientes de (di)similaridad entre las variables y produzcan diferentes configuraciones típicas (*roles de facetas*) espaciales (ver Brown, 1985):

Rol de la faceta	Orden	min. n ^o de elementos	Estructura de la matriz
Axial	Sin orden	2	Simplex
	Linear (independiente)	3	
Conjunta	Linear (dependiente)	3	Circumplex
Modular	Linear (centralidad)	2	
Polar	Sin orden (cualitativa)	3	
	Circular	4	

De esta manera, el SSA permite identificar tanto relaciones lineales como jerárquicas y circulares.

Dado que los distintos aspectos de la actuación docente son categorías cualitativas de un mismo nivel conceptual y, en general, sin orden interno (exceptuando el posible conflicto práctico entre completar el programa y prestar atención a la interacción con los alumnos) esperamos que se organicen en una faceta polar.

4. METODOLOGÍA

Se utilizan dos muestras correspondientes a la Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU) y a la Universidad Autónoma de Madrid (UAM). Se trabaja con un tamaño de muestra de 77.801 para la UPV/EHU y de 90.905 para la UAM.

Los instrumentos aplicados son los conocidos cuestionarios de 'opinión sobre la actuación docente' manejados en estas universidades. Estos cuestionarios son administrados hacia el final de cada cuatrimestre. Su longitud y contenido actual es producto de sucesivas modificaciones y mejora fruto de los análisis estadísticos realizados, del trabajo de comisiones académicas, etc.

El instrumento aplicado en la UPV/EHU es el siguiente:

TABLA 3
CUESTIONARIO APLICADO EN LA UPV/EHU

ítem	Enunciado del ítem
1	Establece, al inicio del curso, el programa y el plan de trabajo de la asignatura
2	Establece, al inicio del curso, los criterios y actividades de evaluación de la asignatura
3	El desarrollo de la asignatura se ha ajustado a lo establecido al inicio del curso
4	El ritmo con que se desarrolla el programa es adecuado
5	Explica con claridad
6	Sus clases están bien preparadas
7	Responde adecuadamente a las preguntas que se le hacen
8	Intenta que el alumno participe en las clases
9	Dialoga con el alumno sobre la marcha de la clase
10	Tiene en cuenta la opinión del alumnado a la hora de decidir las cuestiones relacionadas con la marcha de la asignatura
11	Se esfuerza en que el alumnado se interese por la asignatura
12	Los materiales utilizados y/o recomendados (apuntes, libros...) resultan de utilidad
13	Los ejemplos, prácticas, ejercicios, problemas... que plantea son adecuados en el contexto de la asignatura
14	El sistema de evaluación propuesto es adecuado
15	En general, pienso que es un/a buen/a profesor/a
16	Me siento satisfecho/a asistiendo a sus clases
17	He aprendido mucho cursando esta asignatura
18	El/La profesor/a asiste regularmente a clase
19	En caso de no asistir a clase la recupera
20	En las horas de tutorías está disponible en su despacho
21	Cumple debidamente su horario, comenzando y terminando con puntualidad

Escala de los ítems: 1 «nada de acuerdo»; 2 «poco de acuerdo»; 3 «medianamente de acuerdo»; 4 «bastante de acuerdo»; y 5 «totalmente de acuerdo».

El instrumento aplicado en la UAM es el siguiente:

TABLA 4
CUESTIONARIO APLICADO EN LA UAM

ítem	Enunciado del ítem
1	El/la profesor/a explica con claridad.
2	Se preocupa por que los alumnos aprendan.
3	Suele destacar las cosas que considera importantes.
4	Contribuye a hacer interesante la asignatura.
5	Sus clases están bien preparadas.
6	El/la profesor/a parece dominar la asignatura y estar al corriente de los progresos de la materia.
7	Ha informado sobre el programa y el plan de trabajo de la asignatura.
8	En líneas generales, el/la profesor/a se ha ajustado al plan de trabajo previsto.
9	Ha informado sobre los criterios y actividades de evaluación de la materia que imparte.
10	El/la profesor/a tiene una actitud receptiva ante las preguntas o sugerencias de los estudiantes.
11	Fomenta la participación de los estudiantes en clase.
12	Está disponible para ser consultado/a en horas de tutoría. (contesta sólo en caso de utilizarlas).
13	Los conceptos teóricos se complementan adecuadamente con ejemplos, comentarios de texto, ejercicios, problemas, trabajos, etc.
14	La bibliografía y/o el material de lectura indicados por el /la profesor/a son útiles para el estudio de la asignatura.
15	En general el trabajo llevado a cabo por el/la profesor/a ha sido satisfactorio.

Escala de los ítems: 1 -«totalmente en desacuerdo»; 2-«bastante en desacuerdo»; 3-«en desacuerdo»; 4-«ni de acuerdo ni en desacuerdo»; 5-«de acuerdo»; 6-«bastante de acuerdo»; 7-«totalmente de acuerdo».

Las técnicas estadísticas utilizadas serán, por un lado, el modelo de medida de los modelos de ecuaciones estructurales (Análisis Factorial Confirmatorio AFC) mediante la herramienta AMOS (Arbuckle, 1997) y, por otro, el escalamiento multidimensional no-paramétrico Smallest Space Analysis (SSA - Guttman, 1968).

5. RESULTADOS OBTENIDOS

En primer lugar se realizaron los análisis con Análisis Factorial Confirmatorio mediante la herramienta AMOS. A continuación se realizaron los análisis con la herramienta SSA.

5.1. Análisis factorial confirmatorio (AMOS)

Los resultados de la muestra y el cuestionario de la Universidad del País Vasco se presentan en la Figura 1. Las estimaciones se han realizado mediante el algoritmo ML (Máxima Verosimilitud). Los parámetros estimados de los pesos factoriales en cada uno de los ítems así como los pesos factoriales de la dimensión global sobre los factores de primer orden muestran con claridad la presencia de estructuras de covariación diferenciadas por competencias específicas al mismo tiempo que es evidente la presencia de un factor general potente y común a todos los elementos del instrumento. Los índices de ajuste del modelo se muestran en el propio diagrama y merecen un comentario aclaratorio.

En general, podemos considerar que el ajuste de los datos al modelo es entre moderado y bajo. Sin embargo, esta valoración debe realizarse con matices ya que el nivel de ajuste logrado no puede valorarse en términos absolutos sino en relación a otros modelos alternativos y a las características de los datos. En cuanto a estas últimas hay dos consideraciones importantes que hacer. Por un lado, el gran tamaño de la muestra da lugar a que indicadores tales como el Chi-Cuadrado tengan una probabilidad asociada extremadamente baja. Por otro lado, las correlaciones entre los ítems incluidos en el análisis son bastante altas y es conocido que esto afecta también negativamente a algunos de los indicadores de ajuste. También es preciso señalar que para una valoración adecuada del mayor o menor ajuste de este modelo sería necesario compararlo con modelos alternativos. Sin embargo, este propósito desborda los objetivos del presente trabajo.

Los índices de ajuste del modelo manejados han sido elegidos según dos criterios. Por un lado, presentar los índices más habituales en la investigación del área. Por otro lado, presentar índices de las principales familias o grupos de índices: ajuste absoluto global del modelo, ajuste ponderado por grado de parsimonia. De esta manera, tenemos el Chi-Cuadrado como prueba de contraste de hipótesis, el RMR como promedio de residuales cuadráticos, el GFI como índice absoluto global de ajuste y el PGFI como índice de ajuste relativo a la parsimonia del modelo.

En estos índices, los resultados pueden considerarse suficientes de acuerdo con los criterios recomendados por autoras tales como Byrne (1998) para la cual niveles en GFI superiores a .90, en PGFI superiores a .50 y en RMR inferiores a .05 serían aceptables incluso en modelos con chi-cuadrado no-significativo. En nuestros datos, nos encontramos dentro de estos niveles o muy cercanos a ellos.

Podría haberse logrado una mejora importante en los índices de ajuste mediante la liberación de algunos parámetros guiándose por los denominados 'Índices de Modificación' sin descuidar los aspectos sustantivos. Como en numerosos modelos de medida, los componentes de error de los ítems pueden estar intercorrelacionados por diversas razones. Entre ellas cabe destacar que la formulación de los factores latentes puede no ser exhaustiva y dejar estructuras de covariación residuales no recogidas por la estructura latente planteada. Este hecho tiene una gran relevancia cuando el muestreo de los ítems de cada dimensión o factor es claramente insuficiente como es el caso. Por este motivo se ha prescindido en el análisis del ítem 14 ya que era el único

FIGURA 1
 RESULTADOS ESTANDARIZADOS CON AMOS PARA LA UPV/EHU

chi cuadrado= 33590,171; gl=61; P=,000
 GFI = ,934 PGFI = ,626 RMR = ,056

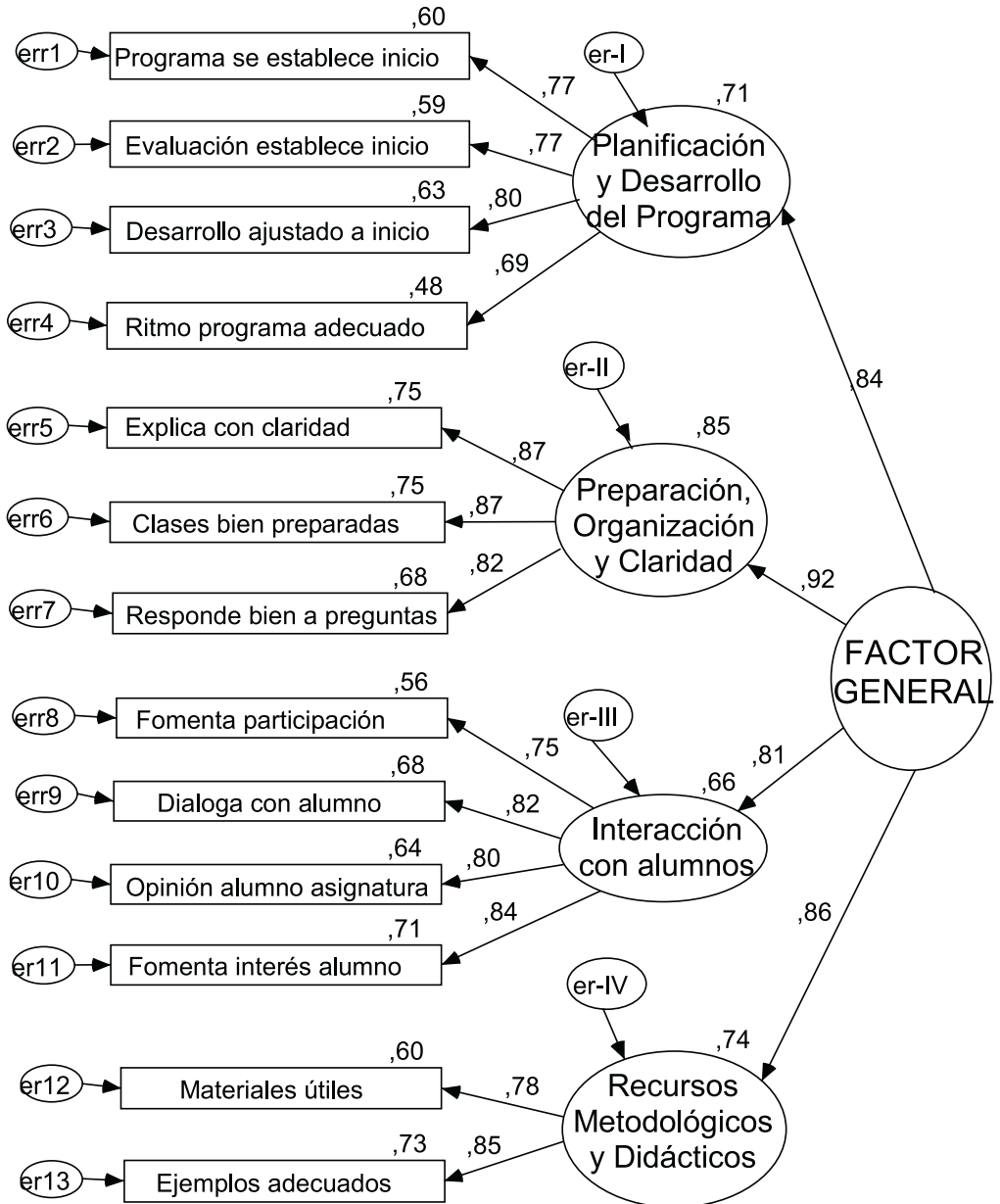
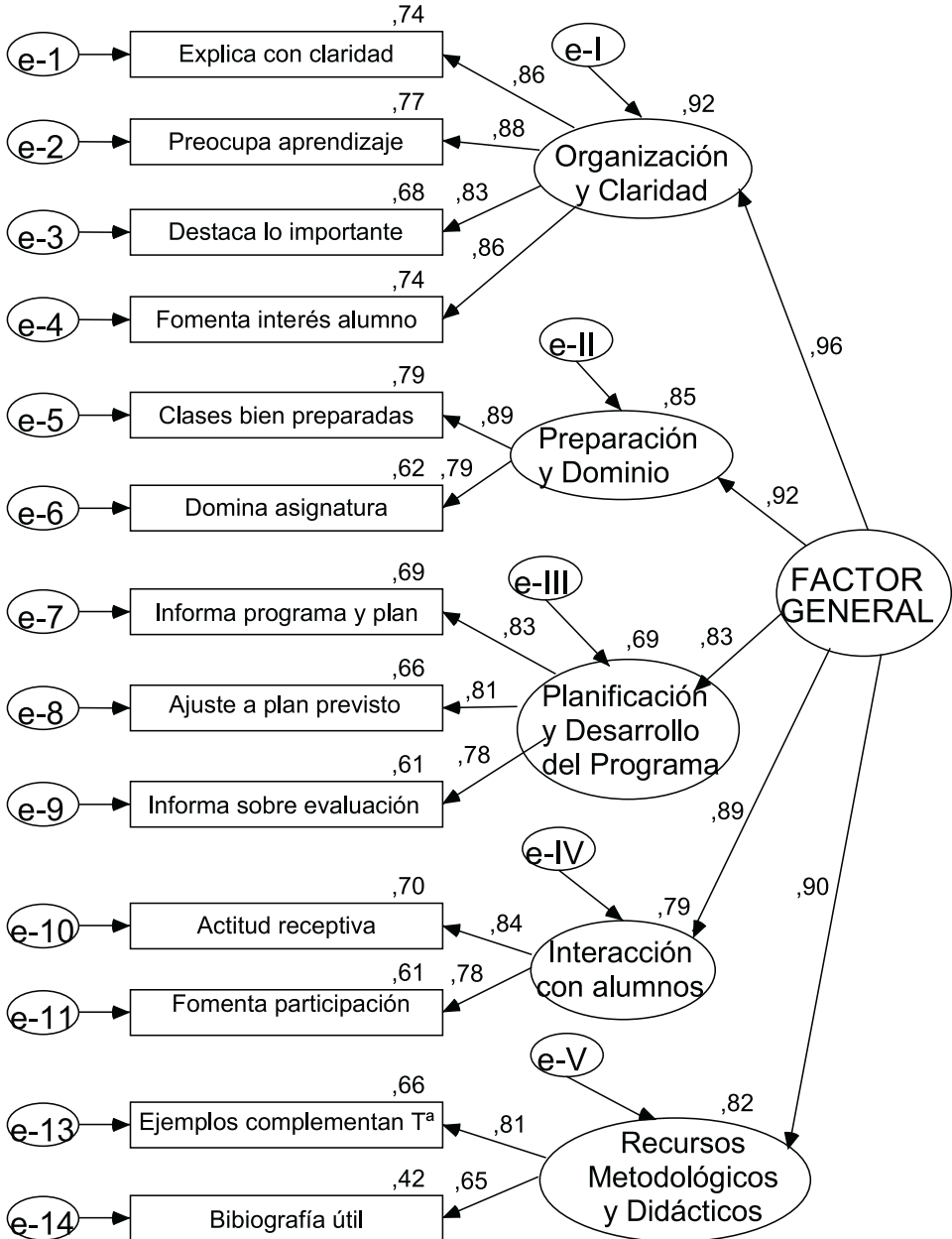


FIGURA 2
 RESULTADOS ESTANDARIZADOS CON AMOS PARA LA UAM

chi cuadrado= 28380,688; gl=60; P=,000
 GFI = ,952 PGFI = ,628 RMR = ,056



del instrumento para un posible factor de 'evaluación'. También hay que considerar la dificultad de utilizar ítems 'puros' explicados por una única dimensión o factor, debido a la necesidad de formular ítems con significado sustantivo para los estudiantes. Por ello, algunos ítems pueden estar explicados por más de un factor. La estrategia de ir liberando parámetros de pesos factoriales para que más de un factor explique el mismo ítem lleva pareja la distorsión de lo característico de cada dimensión. En definitiva, el ajuste del modelo podría mejorar considerablemente pero su consistencia conceptual quedaría mermada. Un ejemplo claro podría ser eliminar de los análisis el ítem número 4 o liberar el parámetro de peso factorial con las dos primeras dimensiones. Este ítem resulta especialmente problemático por cuanto posee componentes de ambas dimensiones.

También hubiera sido factible liberar algunas intercorrelaciones entre los errores dando lugar a unos indicadores de ajuste notablemente superiores. Sin embargo, se ha preferido presentar los resultados de la manera más transparente posible dado que no es el propósito del presente trabajo lograr altos índices de ajuste sino, más bien, hacer evidente empíricamente la presencia simultánea de una multidimensionalidad y de una unidimensionalidad.

Para este propósito, los resultados encontrados en la Figura 1 son clarificadores. Como puede observarse los pesos factoriales de las latentes de primer orden son muy altos con lo que, de alguna manera, queda apoyada la multidimensionalidad del instrumento. Paralelamente, estos factores están muy intercorrelacionados por lo que el planteamiento de un factor general de segundo orden es viable. Como puede observarse, este factor general explica una parte importante de la varianza de los factores de primer orden por lo que queda también apoyada la fuerte unidimensionalidad del instrumento. Así, como interpretación general de los resultados, puede afirmarse la presencia de una estructura jerárquica lineal con la presencia de un factor general potente que, a su vez, se despliega en diversos factores específicos.

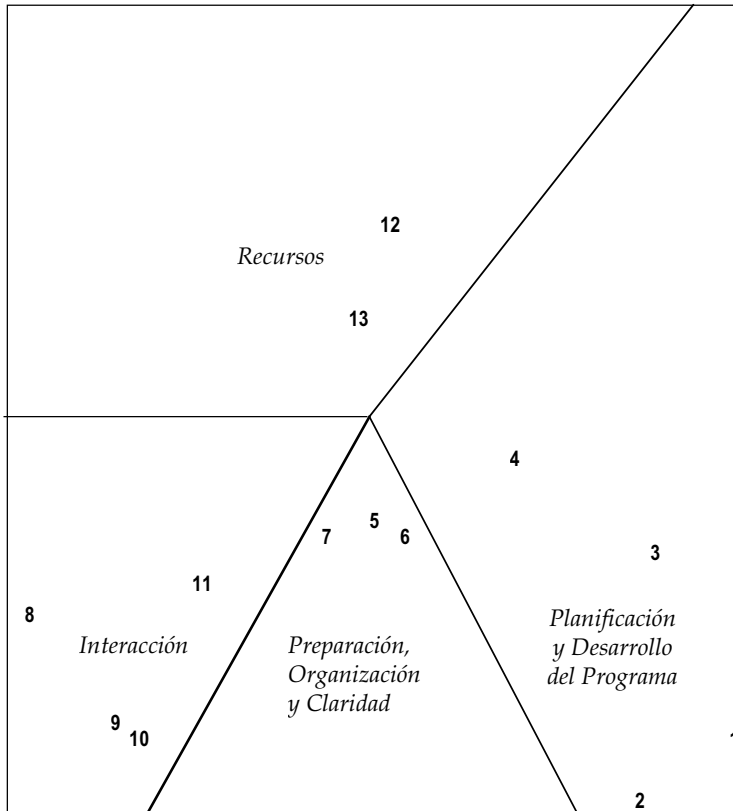
Los resultados encontrados con la muestra y cuestionario de la Universidad Autónoma de Madrid son similares a los encontrados con la muestra de la UPV/EHU. La estructura dimensional es muy parecida si bien existen algunas variantes en la composición de las dimensiones de Dominio y Planificación. Los indicadores de ajuste están en los mismos niveles y cabe aplicar aquí lo mencionado anteriormente. También son evidentes las posibilidades de mejorar sustancialmente el ajuste del modelo mediante ligeras modificaciones que no tienen carácter sustantivo.

Realizando un juicio integrador de los resultados encontrados con ambas muestras y ambos cuestionarios, puede afirmarse que los datos soportan una interpretación jerárquica-lineal del constructo en la cual cabe plantear unas dimensiones específicas de competencia docente que, a su vez, se integran en una dimensión general.

5.2. Análisis estructural (SSA)

Los mismos datos fueron sometidos a SSA bidimensional. El grado de ajuste de la función en ambos casos es aceptable al presentar coeficientes de alienación de .11 y .14. Como se ha indicado anteriormente, podemos considerar que la configu-

FIGURA 3
 RESULTADOS SSA PARA LA UPV. CUESTIONARIO SOBRE DOCENCIA: UPV:
 SSA DE 13 ÍTEMS SOBRE ACTUACIÓN DOCENTE
 2 DIMENSIONES - COEFICIENTE DE ALIENACIÓN: ,11

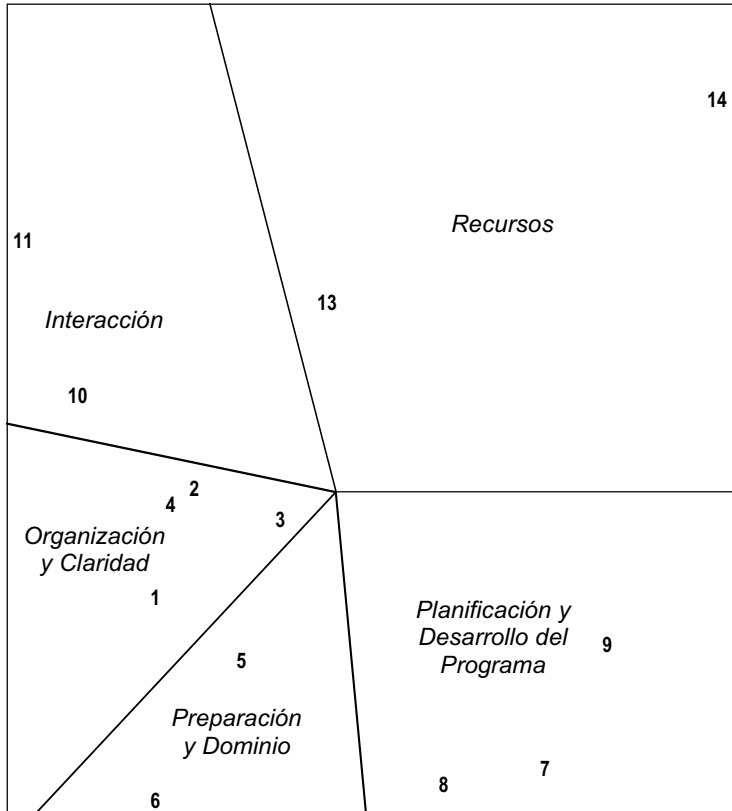


ración presente en el plano refleja con fidelidad los datos al presentar coeficientes de alienación inferiores a .20. Ambas representaciones han podido ser interpretadas conceptualmente identificando los distintos aspectos de la actuación docente presentes en los cuestionarios. En ambos casos, los ítems que reflejan un mismo aspecto de la actuación docente convergen permitiendo delimitar regiones disyuntas.

La aplicación SSA proporciona únicamente la ubicación de los puntos (ítems) en el espacio (sus coordenadas y representación gráfica). Por lo tanto, es el investigador quien delimita las regiones aplicando la teoría previa. Fruto de este análisis se han trazado una serie de líneas divisorias (fronteras entre regiones) sobre el plano que proporciona la aplicación.

De acuerdo a lo esperado, esas regiones producen una configuración polar. Esta configuración está parcialmente ordenada por la contradicción práctica entre la inte-

FIGURA 4
 RESULTADOS DEL SSA PARA LA UAM. CUESTIONARIO SOBRE DOCENCIA: UAM:
 SSA DE 13 ÍTEMS SOBRE ACTUACIÓN DOCENTE
 2 DIMENSIONES - COEFICIENTE DE ALIENACIÓN: ,14



racción (el fomento de la participación, la atención a las opiniones y sugerencias del alumnado) y el desarrollo pleno del programa. Finalmente, la ubicación de los ítems de 'Organización y Claridad' en un sector más central de la representación puede reflejar la centralidad relativa de este aspecto en la percepción de la calidad de la actuación docente por parte de los estudiantes.

6. CONCLUSIONES

La aproximación no paramétrica es menos restrictiva, más sencilla e intuitiva y menos exigente en sus condiciones de aplicación. Se ha hecho evidente la superioridad de esta técnica para mostrar una estructuración del constructo no-jerárquica y una caracterización más precisa de la generalidad y especificidad presentes en cada uno de los ítems. Por lo tanto, la técnica SSA se muestra como un buen heurístico para la

comprensión de fenómenos tales como la competencia docente y para la generación o contraste de teorías sobre la estructura del constructo.

La contraposición entre una orientación unidimensional o multidimensional se presenta como innecesaria al poder contemplar el constructo tanto como unidimensional o multidimensional. En definitiva, serían los usos evaluativos los que definirían la necesidad de utilizar puntuaciones unidimensionales o multidimensionales. El SSA permite verificar los aspectos que organizan la percepción de la actuación docente por parte de los estudiantes. La ubicación de los ítems en las regiones permite, asimismo, identificar aquellos que mejor reflejan cada aspecto. Finalmente, la centralidad relativa en la representación espacial puede ayudar a identificar aquellos ítems que reflejan los aspectos más integradores o globales de la competencia docente según los alumnos.

En este sentido, la ubicación de los ítems que reflejan 'Organización y Claridad' en el centro del plano abre una interesante discusión sobre si estos ítems ('explica con claridad', 'destaca las cosas que considera importantes', 'responde adecuadamente a las preguntas') son indicadores del aspecto más esencial de la competencia docente en opinión de los alumnos. Puesto que este factor es considerado el principal para la eficacia docente también por la teoría didáctica (Braun y Atkins, 1988), el resultado obtenido aportaría evidencia sobre la validez de constructo de los cuestionarios utilizados.

Asimismo, este resultado sugeriría dos opciones alternativas para la estimación de la competencia general de un profesor con fines sumativos. Por un lado, cabría utilizar los ítems de valoración general ('en general, el trabajo llevado a cabo por el profesor ha sido satisfactorio', '...', 'pienso que es un buen profesor', 'pienso que he aprendido mucho'), mientras que, por otro lado, cabría utilizar los ítems referidos a conductas o competencias más concretas, de dominio y claridad en la explicación, que la teoría didáctica considera y los análisis revelan como esenciales para la eficacia docente.

La segunda alternativa podría tener efectos positivos pero también perversos. Por el lado positivo, resulta evidente que el profesorado que recibe feed-back sobre conductas concretas o específicas estará mejor orientado para mejorar su práctica en dichos aspectos. Por el contrario, si recibe un feed-back genérico sobre la valoración o satisfacción general de sus alumnos, sus pautas de actuación/mejora serán más imprecisas.

Por el lado de los efectos perversos cabría subrayar que el profesorado podría percibir que son incentivadas las conductas más relacionadas con una docencia magistral o tradicional, en detrimento de otras conductas o competencias más específicas e innovadoras como podrían ser las referidas al aspecto de interacción. Sin embargo, los ítems referidos a 'Organización y Claridad' parecen también ser también más inmunes a la influencia de factores contextuales (en comparación, por ejemplo, con el efecto del tamaño de grupo sobre los ítems referidos a la interacción con los alumnos). De alguna manera, estos ítems reflejarían una cierta 'generalidad' desde varios puntos de vista.

En definitiva, cabe recomendar el uso del análisis estructural no-paramétrico como complementario al clásico AFC con el fin de identificar con mayor claridad la estructura del constructo y la relativa centralidad-generalidad de cada uno de los ítems.

Mientras los usos sumativos de estas evaluaciones seguirán requiriendo ítems de valoración general, los usos formativos requerirán la estimación de la competencia en las dimensiones específicas de la actuación docente-poniendo especial atención a la centralidad relativa del aspecto de 'Organización y Claridad' desde los puntos de vista empírico y teórico. Los elementos de este aspecto de la actuación docente pueden constituir el punto de apoyo o anclaje sobre el que desarrollar las competencias docentes específicas.

7. REFERENCIAS

- Abrami, P. C. (1989). How should we use student ratings to evaluate teaching? *Research in Higher Education*, 30(2), 221-227.
- Abrami, P. C., & Apollonia, S. (1990). The Dimensionality of Ratings and Their Use in Personnel Decisions. In M. Theall & J. Franklin (Eds.), *Student Ratings of Instruction: Issues for Improving Practice* (pp. 97-112). San Francisco: Jossey-Bass.
- Abrami, P. C., & Apollonia, S. (1999). Current concerns are past concerns. *American Psychologist*, 54(7), 519-520.
- Aparicio, J. J., Tejedor, J., & Sanmartín, R. (1982). *La enseñanza universitaria vista por los alumnos: Un estudio para la evaluación de los cursos en la enseñanza superior*. Universidad Autónoma de Madrid. Instituto de Ciencias de la Educación.
- Apodaca, P. (1995). *Cuestionario de docencia teórica: estudio métrico-sustantivo*. Unidad de Evaluación: Universidad del País Vasco. Informe no publicado.
- Apodaca, P. y Páez, D. (1992). Los modelos estructurales, posibilidad y limitaciones. En D. Páez, J. Valencia, J.F. Morales, B. Sarabia y N. Ursúa (Eds.): *Teoría y Método en Psicología Social*. Madrid: Anthropos.
- Apodaca, P. y Rodríguez, M. (1999). La opinión de los alumnos en la evaluación de la calidad docente: posibilidades, limitaciones y estructura dimensional de sus indicadores. En J. Vidal (Coord.): *Indicadores en la Universidad: información y decisiones*. Madrid: Consejo de universidades, Ministerio de Educación y Cultura.
- Apollonia, S. & Abrami, P. (1997). Navigating Student Ratings of Instruction. *American Psychologist*, 52, 11, 1198-1208.
- Arbuckle, J.L. y Wothke, W. (1999). *Amos User's Guide: Versión 4.0*. Chicago: SPSS.
- Batista, J.M. y Coenders, G. (2000). *Modelos de Ecuaciones Estructurales*. Madrid: La Muralla.
- Borg, I. (1981). *Multivariate data representations: when and why*. Ann Arbor, MI: Mathesis Press.
- Borg, I., & Groenen, P.J.F. (1997). *Modern Multidimensional Scaling: Theory and applications*. New York: Springer.
- Borg, I., & Lingoes, J. (1987). *Multidimensional Similarity Structure Analysis*. New York: Springer-Verlag.
- Borg, I., & Shye, S. (1995). *Facet theory: Form and content*. Thousand Oaks, CA: SAGE (Advanced Quantitative Techniques in the Social Sciences series, Vol. 5).
- Braun, G. & Atkins, M. (1988). *Effective teaching in Higher Education*. London: Routledge.

- Brown, J. (1985). An introduction to the uses of facet theory. En D. Canter (Ed.). *Facet theory: Approaches to social research* (pp. 17-57). New York: Springer-Verlag.
- Byrne, B.M. (1998). *Structural Equation Modeling with LISREL, PRELIS and SIMPLIS: Basics Concepts, Applications, and Programming*. London: Lawrence Erlbaum Associates.
- Canter, D. (Ed.). (1985). *Facet theory: Approaches to social research*. New York: Springer-Verlag.
- Cohen, P. A. (1981). Student ratings of instruction and student achievement: a meta-analysis of multisection validity studies. *Review of Educational Research*, 51(3), 281-309.
- Dancer, L.S. (1990). Introduction to facet theory and its applications. *Applied Psychology: An international review*, 39, 365-377.
- Davison, H. (1983). *Multidimensional Scaling*. New York: Wiley.
- Davison, M.L. (1985). Multidimensional scaling versus component analysis of test intercorrelations. *Psychological Bulletin*, 97, 94-105.
- De Miguel, M. (1998). La evaluación del profesorado universitario. Criterios y propuestas para mejorar la función docente. *Revista de Educación*, 315, 67-84.
- De Miguel, M., Arias, J. M., Fernández-Raigoso, M., Fueyo, A., & Quiros, J. C. (1991). *Criterios para la evaluación del profesorado universitario*. Oviedo: Universidad de Oviedo.
- Dillon, W.R. y Goldstein, M. (1984). *Multivariate analysis: Methods and approaches*. New York: Wiley.
- Elejabarrieta, F. y Perera, S. (1989). Las paradojas de los rigores metodológicos en psicología social. *Revista Vasca de Psicología*, 2, 1, 81-90.
- Feldman, K. A. (1989). Instructional effectiveness of college teachers themselves, current and former students, colleagues, administrators, and external (neutral) observers. *Research in Higher Education*, 30(2), 137-193.
- Feldman, K. A. (1996). Identifying exemplary teaching: using data from course and teacher evaluations. *New Directions for Teaching and Learning*(65), 41-50.
- García Ramos, J. M. (1997). Análisis Factorial Confirmatorio en la validación del constructo competencia docente del profesor universitario. *Bordón*, 49, 4, 361-391.
- Guttman, L. (1968). A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika*, 33, 469-506.
- Guttman, R., & Greenbaum, C.W. (1998). Facet theory: Its development and current status. *European Psychologist*, 3, 13-36.
- Levy, S. (Ed.) (1994). *Louis Guttman on theory and methodology: Selected writings*. Dartmouth Benchbark.
- Lingoes, J.C., & Guttman, L. (1967). Nonmetric factor analysis: a rank-reducing alternative to linear factor analysis. *Multivariate Behavioral Research*, 2, 485-505.
- Marsh, H. W. (1988). Student evaluations of teaching. In M. J. Dunkin (Ed.), *The international encyclopedia of teaching and teacher education* (2nd ed., pp. 181-187). Oxford: Pergamon Press.
- Marsh, H. W., & Dunkin, M. J. (1992). Students' evaluations of University Teaching: A multidimensional perspective, *Higher Education: Handbook of theory and research* (Vol. VIII, pp. 143-233).

- Marsh, H. W., & Hocevar, D. (1984). The factorial Invariance of Student Evaluations of College Teaching. *American Educational Research Journal*, 21(2), 341-366.
- Marsh, H. W., & Hocevar, D. (1991). The multidimensionality of students' evaluation of teaching effectiveness: The generality of factor structures across academic discipline, instructor level, and course level. *Teaching & Teacher Education*, 7(1), 9-18.
- Marsh, H. W., & Roche, L.A. (1997). Making Students' Evaluations of Teaching Effectiveness Effective. *American Psychologist*, 52 (11), 1187-1197.
- Maruyama, G.M. (1998). *Basics of Structural Equation Modeling*. Thousand Oaks, California: Sage.
- Mateo, J. (1987). La evaluación del profesorado universitario: Algunas reflexiones en torno al estado de la cuestión. In M. J. Carrera Gonzalo (Ed.), *Consideraciones Metodológicas sobre la Evaluación y Mejora de la Docencia Universitaria* (pp. 13-30). Valencia: Universidad de Valencia.
- McKeachie, W.J. (1997). Student Ratings: The Validity of Use.. *American Psychologist*, 52, 11, 1218-1225.
- Satorra, A. y Stronkhorst, L.H. (1984). Introducción a los modelos de causalidad. En J.J. Sánchez Carrión (ed.): *Introducción a las técnicas de análisis multivariante aplicadas a las ciencias sociales*. Madrid: CIS.
- Shye, S. (1988). Inductive and deductive reasoning: A structural reanalysis of ability tests. *Journal of Applied Psychology*, 73, 308-311.
- Shye, S., & Elizur, D. with Hoffman, M. (1994). *Introduction to Facet Theory: Content design and intrinsic data analysis in behavioral research*. Thousand Oaks, CA: SAGE (Applied Social Research Methods Series, Vol. 35).
- Tejedor, F. J. (1993). *Experiencias españolas de evaluación de la enseñanza universitaria y nuevas perspectivas*. III Jornadas nacionales de Didáctica Universitaria. Evaluación y desarrollo profesional. Universidad de Las Palmas de Gran Canaria, Servicio de Publicaciones. Las Palmas, Gran Canaria.
- Tejedor, F. J., Jato, E., & Mínguez, C. (1987). Evaluación del Profesorado Universitario por los Alumnos en la Universidad de Santiago. In M. J. Carrera Gonzalo (Eds.), *Primeras Jornadas Nacionales sobre Evaluación y Mejora de la Docencia Universitaria* (pp. 133-210). Valencia: Universidad de Valencia.
- Tejedor, F. J., Jato, E., & Mínguez, C. (1988). Evaluación del profesorado universitario por los alumnos de la universidad de Santiago. *Studia Paedagogica* (20), 73-134.
- Touron, J. (1989). La validación de constructo: su aplicación al CEED (cuestionario de evaluación de la eficacia docente). *Bordón*, 1989, 735-756.
- Villa, A. y Morales, P. (1993). *La evaluación del profesor*. Vitoria-Gasteiz: Servicio Central de Publicaciones del Gobierno Vasco.