

DIAGNÓSTICO EN LA TELEFORMACIÓN: CONSTRUCCIÓN Y VALIDACIÓN DE UN ESCALOGRAMA GUTTMAN

Rafael García Pérez*

Universidad de Sevilla

RESUMEN

El estudio metodológico que presentamos en este artículo tiene por objeto la construcción y validación de un escalograma para instrumentalizar algunas funciones diagnósticas en la teleformación de estudiantes universitarios. También, en un plano más conceptual, planteamos la idoneidad de los escalogramas Guttman como instrumentos de evaluación diagnóstica en este tipo de procesos educativos. Se revisan algunas de las propiedades metodológicas que hacen de los escalogramas herramientas muy apropiadas para ser integradas en el diseño de las páginas Web educativas. Este tipo de instrumento de diagnóstico permite resolver de forma eficaz la autoevaluación inicial en los sistemas de teleformación; especialmente, cuando están diseñados para atender a una diversidad de usuarios con diferentes niveles formativos. Este es el caso del S.T. MIDE, creado como apoyo virtual a la docencia universitaria para el aprendizaje de la materia «Técnicas e Instrumentos de Diagnóstico». Se presenta un escalograma que los usuarios pueden utilizar para la autoevaluación inicial de su nivel de dominio en dicha materia, diferenciando entre niveles de «principiante», «iniciado» y «experto» para la autoselección de actividades de aprendizaje. En suma, se presenta un estudio empírico de la fiabilidad y validez (de constructo y concurrente) del escalograma construido para esta función diagnóstica en la teleformación universitaria.

Palabras claves: Escalogramas Guttman, Técnicas e Instrumentos de Diagnóstico, Teleformación, Sociedad del Conocimiento, Fiabilidad y Validez, Validez de Constructo, Validez Concurrente.

* RafaelGarcia@us.es

ABSTRACT

The methodological study that we present in this article it considers its object the construction and validation of a scalogram to explore some functions of assessment in the learning of university students with systems of «e-learning». Also, in a more conceptual line, we present the Guttman's scales as instruments of assessment to study this type of educational processes. It's revised some methodological properties of the scalogram, that it are very appropriate tools for to be integrated in the design of the educational Web pages. This assessment instrument permits, to the users, resolve of efficient form for the initial evaluation of themselves in the systems of «e-learning»; especially, when it's designed to attend to a diversity of users with different educational levels. This is the case of the «S. T. MIDE», which it's created like virtual support to the learning of the subject: «Techniques and Instruments of Assessment». It's presented an scalogram that the users can utilize for the initial evaluation of themselves, of its level of control in subject, differentiating among levels of «inexperienced», «initiated» and «expert» for the auto-selection of activities of e-learning. In short, it's presented an empirical study of the reliability and validity (of construct and concurrent) of the scalogram elaborated for the assessment function in the university e-learning.

Key words: *Guttman's Scales/ Scalogram, Assessment Techniques and Procedures, Web-based Teaching/E-Learning, Knowledge Society, Reliability and Validity, Construct Validity, Concurrent Validity.*

I. INTRODUCCIÓN

En los últimos cursos académicos hemos trabajado en el diseño de nuevos modelos didácticos para articular la enseñanza presencial de forma combinada con la virtual, en referencia al concepto emergente «blended-learning» (Cornella, 2002); utilizando, como vía de desarrollo profesional, las distintas convocatorias de proyectos de innovación propuestas por el ICE de la U. de Sevilla.

Las conclusiones más genéricas de estos procesos de mejora de nuestra enseñanza pueden consultarse en los informes publicados (García y Otros, 2001; García, 2002). Dichos proyectos de innovación e investigación en la enseñanza universitaria se vinculan a la formación del pedagogo en la Sociedad del Conocimiento; algo especialmente importante hoy en Andalucía, dado que nuestra Comunidad Autónoma pretende una segunda modernización basada en la incorporación de nuestro tejido empresarial y tecnológico a la Sociedad del Conocimiento (Castaño y Román, 2002). En este sentido, la voluntad expresa de nuestra comunidad de incorporarse a la «Sociedad Red», como vía de superación de algunos de nuestros problemas históricos (retraso social y tecnológico, paro, ...), queda recogido en la serie de documentales «El Siglo de las Luces», elaborados y emitidos por Canal Sur bajo la dirección de Manuel Campo Vidal y la colaboración expresa de Manuel Castells, en un convenio que vincula a distintas cadenas televisivas autonómicas (Canal Nou, Televisión Galega, ...).

En nuestro proyecto actual, se han realizado labores de creación y evaluación de recursos didácticos cara a la formación técnico-profesional y científica en el marco de la

materia «Técnicas e Instrumentos de Diagnóstico (TID)», lo que se traduce actualmente en la disponibilidad de una serie de entornos y actividades de aprendizaje virtual diseñadas ad hoc para el aprendizaje de los contenidos de la asignatura. Todo ello, forma parte del Sistema de Teleformación de apoyo a la docencia presencial en el área MIDE (<http://www.us.es/lablic/STMIDE/STMIDE.htm>). Este se viene construyendo, de forma progresiva y con diferentes apoyos institucionales, en el laboratorio L.I.C. desde su creación en 1998/99. Precisamente, el Laboratorio de experimentación e Investigación educativa de procesos Culturales (L.I.C), dirigido por la Dra. Pilar Colás Bravo, se crea con el interés explícito de dar cobertura a todo un amplio abanico de nuevas prácticas docentes y sistemas de formación en el área MIDE del Dpto. DOE y MIDE de la Universidad de Sevilla.

El propósito de estos sistemas es complementar la enseñanza presencial tradicional con otros mecanismos cara a «fomentar las capacidades críticas y creativas que permitan al estudiante buscar, recuperar, organizar y desarrollar dicha información como conocimiento personalizado que pueda serle útil en su posterior vida social y profesional» (González García y otros, 1996: 233). Según la experiencia que hemos acumulado en este laboratorio, lo esencial cara a la construcción y valoración de dichos sistemas es la propuesta educativa que se implementa con ellos. En este sentido, las propuestas pedagógicas «innovadoras» que hemos realizado se basan en dos procesos: *el trabajo colaborativo en grupos de alumnos y la resolución de problemas*. Ambos procesos se articulan en un espacio más «horizontal» de colaboración entre iguales (alumnos) para promocionar la responsabilidad de la enseñanza recíproca de los estudiantes hacia sus propios compañeros utilizando las nuevas tecnologías de aprendizaje en Internet.

El alumno juega un papel central en este modelo didáctico; pues todo ello, implica considerar a los estudiantes desde una idea de excelencia y exigencia dado que su responsabilidad se amplifica en el marco de actividad que supone para los alumnos la *enseñanza recíproca*, de manera que los resultados de su trabajo influyen directamente en la formación y aprovechamiento para el aprendizaje de sus propios compañeros. Ello, también les supone la vivencia personal de un conjunto de «experiencias particulares» sobre la formación y el diagnóstico en la Sociedad del Conocimiento; así como un replanteamiento de las relaciones horizontales con sus compañeros en la asignatura.

Concretamente, los alumnos deben construir colaborativamente en grupos páginas web «educativas», resolviendo problemas de adaptación de los contenidos al lenguaje multimedia en HTML, que están dirigidas al aprendizaje de una técnica o instrumento concreto para el diagnóstico educativo. Además deben realizar experiencias de aprendizaje con las páginas elaboradas por los compañeros y evaluarlas según su capacidad para regular y ayudar en dichos aprendizajes.

Son precisamente estas peculiaridades del modelo didáctico propuesto las que justifican su experimentación ante los procesos de alfabetización digital y científica que requieren hoy los estudiantes universitarios que se preparan para ejercer profesionalmente el diagnóstico educativo en la Sociedad del Conocimiento. La incertidumbre ante la inmensa cantidad de información que disponemos hoy en las redes informáticas, la capacidad crítica ante las fuentes de información en función de su fiabilidad y aplicabilidad a la resolución de problemas científicos concretos y la flexibilidad en la

integración de propuestas elaboradas desde perspectivas muy diversas, constituyen rasgos positivos de este nuevo marco de aprendizaje. Los alumnos se ven inmersos en dicho entorno, en el que deben desarrollar, ampliar y cuestionar críticamente sus conocimientos.

El S. T. MIDE, en un plano metodológico-didáctico, recoge y organiza 432 actividades didácticas sobre los contenidos de la asignatura. Estas recorren un amplio arco de posibilidades de aprendizaje, considerando distintos tipos de métodos de formación según el nivel de dominio y conocimiento que tienen los usuarios sobre las «técnicas e instrumentos de diagnóstico» (principiante, iniciado y experto). El tipo de actividades de aprendizaje, que se recogen en el sistema de teleformación, se formulan en base a una combinación de los cuatro métodos didácticos que constan a continuación; siendo los dos primeros más utilizados en los niveles de principiante e iniciado y los dos segundos más propios de los niveles de formación de expertos:

- 1- **Actividades de estudio independiente:** Para aprender cómo pueden usar los recursos del sistema y establecer pautas de orientación, ... (p.e.: lectura de libros, artículos, contenidos de blended-learning, etc.) para el aprendizaje.
- 2- **Actividades de manipulación práctica:** Dirigidas al dominio instrumental y al desarrollo de trabajos tanto individuales como colectivos (Ej.: elaboración y aplicación de un test para un examen, una escala de actitudes, etc.).
- 3- **Actividades de discusión:** Se especifica el tipo de temática que interesa discutir, qué modelos de grupos (grandes, pequeños, symposium...) y la organización dinámica del aprendizaje (mediante foros en Internet, BCSW...).
- 4- **Actividades de investigación y de acción profesional:** (p.e.: investigar sobre algún problema que plantean los instrumentos; o, sobre cualquier otra temática que tenga relación con la práctica profesional del mismo).

Cada grupo de alumnos o usuarios puede elaborar su parrilla didáctica y realizarla. Cada actividad que compone el sistema de teleformación se presenta con una guía explicativa incorporando los distintos elementos didácticos, especificando en cada actividad los siguientes aspectos:

- ⇒ Identificación de objetivos y contenidos que se conjugan con cada actividad.
- ⇒ Explicación del objetivo principal de cada actividad y fases en su realización.
- ⇒ Disposición y propuesta de los recursos para hacer la actividad (en Internet).
- ⇒ Propuesta de otras aplicaciones o ejemplos similares de uso del instrumento.
- ⇒ Propuesta de sistemas de autoevaluación en Internet sobre el propio proceso de innovación y el aprendizaje.

Los sistemas de evaluación, que nos sirven tanto para el seguimiento del alumno como del proyecto de innovación, se exponen a continuación. Por otra parte, la evaluación del modelo didáctico aplicado ha sido objeto de proyectos de innovación anterior apoyados por el ICE-US, cuyos resultados constan en la bibliografía citada (García y otros, 2001; García, 2002).

2. DIAGNÓSTICO EN LA TELEFORMACIÓN: FUNCIONES, MOMENTOS, TÉCNICAS E INSTRUMENTOS

El diagnóstico de esta experiencia de innovación, entendido globalmente como proceso de «assessment», consta de una cierta complejidad en su arquitectura didáctica. Hemos aplicado muy diversas técnicas e instrumentos para constatar datos que asumen diversas funciones y objetivos de diagnóstico; ello, en los diversos «momentos» clave de la evaluación del conjunto de iniciativas de teleformación que conforman el programa de innovación antes expuesto. En un resumen muy escueto son los siguientes:

- *Evaluación diagnóstica*: Para la correcta selección de la propia parrilla didáctica de actividades, según la propia necesidad e interés de aprendizaje en el ámbito de cada técnica diagnóstica (autoevaluación del nivel —principiante, iniciado y experto—) mediante el Escalograma Guttman autoaplicable descrito en este artículo).
- *Evaluación del diseño* del entorno virtual de aprendizaje dispuesto para cada técnica e instrumento de diagnóstico. Se realiza mediante una lista de control y una escala Likert sobre la calidad del diseño de los contenidos y de la organización y explotación didáctica de las posibilidades expresivas e interactivas en Internet.
- *Evaluación del proceso*: Para constatar las dificultades de aprendizaje en el proceso de teleformación, así como la adecuación del sistema para regular el proceso de desarrollo educativo de los alumnos. Se realiza mediante cuestionario a través de Internet. Asimismo, un foro-web de «*seguimiento tutorial* entre pares», permite a cada grupo de trabajo, como responsable de sus contenidos de e-learning, coordinar los procesos de aprendizaje y ofrecer apoyo virtual al resto de grupos del aula.
- *Evaluación del producto*: Para constatar en qué grado se van consiguiendo los progresos en el aprendizaje que se proponen con el sistema de teleformación. Esta evaluación global se articula mediante el test adaptativo informatizado y autoaplicable «TESTAITID», elaborado en el Laboratorio L.I.C. para esta innovación.

El planteamiento de este estudio se restringe particularmente a la evaluación inicial o diagnóstica antes señalada, desarrollando expresamente el escalograma como un instrumento que cubre una serie de objetivos para el funcionamiento del S.T. MIDE.

3. IDONEIDAD DE LOS ESCALOGRAMAS GUTTMAN EN LA TELEFORMACIÓN

Nuestra propuesta postula una idea, quizás novedosa, que vincula el desarrollo interdisciplinar del Área MIDE con el ámbito de la Tecnología Educativa (TE/NTIC, que es también parte de nuestra identidad como grupo de investigación). Quizás por la confluencia en un mismo equipo, el Grupo de Investigación, Evaluación y Tecnología Educativa (P.A.I.: HUM-154), de varias perspectivas sobre el papel de la ciencia y

la tecnología en la sociedad, hoy nos parece indiscutible el impacto de las TICs sobre la investigación científica y sobre la construcción del conocimiento científico (Colás, 2002). Tanto es así que, desde mediados de la década de los noventa, venimos defendiendo que debemos tener un espacio de discusión y desarrollo común como enseñantes de la acción (científica/educativa) llevada a cabo con tecnologías. Aunque aún está en ciernes, la asociación nacional de profesorado universitario SIDITE y AIDIPE podrían contribuir decisivamente a constituir de ese espacio comunicativo.

En este marco de reflexión descubrimos que la nueva didáctica tecnológica que nos propone la Sociedad del Conocimiento, no sólo abre nuevas vías en el campo de las técnicas de investigación educativa, sino que incluso recupera vías que, en alguna medida, ya se consideran superadas por otras para determinadas funciones diagnósticas. Este parece el caso de las técnicas basadas en los «Escalogramas Guttman» (Guttman, 1950; 1984); un tipo de escala acumulativa de medida ordinal, aplicada originalmente a las actitudes y que, en cierta forma, fue desplazada por las técnicas de escalamiento, más sencillas de elaborar, propuestas por Likert (Summer, 1984; Morales, 2000).

Los escalogramas Guttman conllevan una elaboración compleja que implica una cierta formación metodológica. Sin embargo, nuestra experiencia de innovación mediante la creación de un sistema virtual de apoyo a la docencia presencial, pone de manifiesto que los escalogramas tienen unas propiedades que los hacen especialmente idóneos en la Sociedad del Conocimiento. Este instrumento tiene la propiedad de ser de muy sencilla autoaplicación y muy indicativo cuando se requiere una evaluación diagnóstica rápida y sin gran esfuerzo. Además, aunque su construcción es compleja, no es inaccesible, como podrán comprobar en nuestro ejemplo. Antes de las TICs, el modelo didáctico general se focalizaba sobre el profesor; y en dicho contexto, las herramientas docentes y su facilidad de elaboración y uso según las competencias docentes era lo fundamental. En el nuevo contexto didáctico-tecnológico actual, focalizado sobre el aprendiz, donde el profesor juega un papel más secundario, como mediador y facilitador, es más idóneo cualquier instrumento técnico por su facilidad de uso antes que por su facilidad de construcción.

Cuando un usuario (entiéndase estudiante) accede a un sistema de teleformación, que está diseñado contemplando diversos niveles de profundidad y dificultad de uso, no puede ser sometido a un proceso diagnóstico complejo que agote el esfuerzo aprendizaje que el sujeto pone en juego. Un sistema costoso como vía de diagnóstico; por ejemplo, los tests adaptativos informatizados, suponen al sujeto una situación «de prueba» mucho más costosa en esfuerzo mental y tiempo de atención ante la pantalla que un escalograma. Ello puede contribuir a malgastar los esfuerzos de alumnado en la preparación de las materias, al estudio independiente y a la colaboración en equipos de trabajo (virtuales). Pensamos que es mejor proponer actividades «de prueba», tales como los exámenes con TAIs, cuando se pretende la evaluación de los progresos y los productos en el desarrollo educativo. En dicho «momento de la evaluación del producto del aprendizaje», los TAIs también son un ejemplo de como la tecnología revoluciona el campo de la medida y la evaluación educativa, disfrutando de todos los argumentos que ahora exponemos sobre los escalogramas.

Todo ello supone una *reapropiación*, en el ámbito de los métodos y técnicas de investigación, diagnóstico y evaluación, *de antiguos contenidos y también la creación*

de nuevos sistemas y métodos de diagnóstico. En parte, esto es lo que creemos ocurre hoy con las técnicas basadas en el escalograma; las cuales, al cambiar el contexto tecnológico de la enseñanza, vuelven a ser metodológicamente competitivas según sus propiedades de uso. Igualmente se puede ejemplificar con el caso de los tests, en alguna medida devaluados y didácticamente criticados en sus primeras versiones clásicas, que ahora vuelven de la mano de la Teoría de Respuesta a los Ítems (TRI) y la informática, convertidos en instrumentos adaptativos, más precisos, rápidos y cómodos de usar.

4. UN ESCALOGRAMA GUTTMAN PARA DIAGNÓSTICAR EL NIVEL DE DOMINIO (PRINCIPIANTE, INICIADO Y EXPERTO) DE LAS «TÉCNICAS E INSTRUMENTOS DE DIAGNÓSTICO (TID)»

Un escalograma hace referencia a una especie de escalera conceptualmente bien construida en la que cada escalón supone un paso necesario para el siguiente. Diversos autores (Sierra Bravo, 1992; Ander-Egg, 1996; Buendía, 1998), coinciden en poner ejemplos muy didácticos respecto a los escalogramas; el más extendido se refiere a la evaluación del nivel de alfabetización en el cálculo aritmético para individuos en niveles iniciales de la enseñanza: el primer escalón sería preguntar si se sabe sumar, la respuesta positiva es imprescindible para que el sujeto responda también positivamente a la cuestión de si sabe multiplicar y ésta, es igualmente necesaria respecto de si sabe dividir. En un nivel tan básico parece fácil construir un escalograma, pero no lo es tanto en el caso de las materias que impartimos en la universidad.

4.1. Naturaleza y objetivo del escalograma

El objetivo del escalograma en el marco de la innovación presentada es servir como instrumento de diagnóstico cuando un usuario accede al sistema de teleformación S. T. MIDE. Dicho instrumento debe servir para que el sujeto se autoevalúe, de forma sencilla y casi sin inversión de esfuerzo mental, sobre su conocimiento respecto a cualquier técnica e instrumento de diagnóstico (p.e. los test, las historias de vida, los grupos de discusión, las técnicas proyectivas, ...). El escalograma se diseña para posibilitar una evaluación que permita al usuario discriminar si se encuentra en un nivel de conocimiento y dominio (rasgo medido) correspondiente a «principiante», «iniciado» o «experto». Ello, como paso previo para que pueda seleccionar actividades didácticas «a su medida» sobre el contenido que debe aprender en el sistema; o bien, desarrollar una autoevaluación sobre muy diversas técnicas con el objeto de construirse una parrilla de actividades didácticas autoseleccionadas del sistema de teleformación. El arco de contenidos que debe abarcar este escalograma hace referencia global tanto a objetivos conceptuales como procedimentales y actitudinales de la formación en la materia.

Una vez contextualizado el interés del Escalograma Guttman, desarrollamos a continuación su presentación en la versión final y el proceso de elaboración que ha conducido a dicha versión para que pueda ejercer correctamente su función diagnóstica.

4.2. El Escalograma Guttman sobre dominio de las TID

Combinando todos los requisitos expuestos sobre la naturaleza y funciones del escalograma, que requiere el diseño del S.T. MIDE, hemos llegado a la elaboración del instrumento autoaplicable que consta en la tabla 1 siguiente:

TABLA 1
ESCALOGRAMA (PRINCIPIANTES, INICIADOS, EXPERTOS) EN TID

Escalograma para evaluar el nivel de dominio de las técnicas e instrumentos de diagnóstico en educación (principiante, iniciado y experto).		
ÍTEMS QUE COMPONEN LA ESCALA (CR= 0.9098)	Si	No
1. Usted conoce este instrumento o estrategia de diagnóstico y evaluación educativa	1	0
2. Usted sabe aplicar esta técnica de diagnóstico en un caso o situación concreta	1	0
3. Usted puede identificar y explicar los elementos/pasos fundamentales que la componen	1	0
4. Usted puede elaborar un material educativo sobre su naturaleza y uso en educación	1	0
5. Usted conoce las teorías que fundamentan su aplicación en el campo educativo	1	0
6. Usted podría formar al profesorado o a técnicos sobre el uso de esa técnica en educación	1	0
7. Usted puede plantear argumentaciones técnicas y científicas sobre su uso en educación	1	0
8. Usted conoce las líneas de investigación y últimos avances sobre los enfoques y teorías científicas que fundamentan ésta técnica	1	0
* Nota para su uso: Responda positiva o negativamente a las preguntas aplicándolas a la técnica e instrumento de diagnóstico cuyo conocimiento usted de desee autoevaluar. <u>El primer escalón respondido negativamente indica el nivel del usuario</u> ; los dos primeros escalones corresponden con el nivel de «principiante», los escalones 3, 4 y 5 son del nivel de «iniciado» y los ítems 6, 7 y 8 corresponden a una problemática educativa vinculada al nivel de «experto».		

En un apartado ulterior se expone el estudio empírico que nos ha llevado a considerar la forma presentada para el instrumento definitivo y sus instrucciones de uso. Hay que tener en cuenta que es necesario reconocer la adecuación y calidad de ambos aspectos dado que la aplicación del mismo la ejercen los propios usuarios.

4.3. Proceso metodológico aplicado al estudio piloto de este instrumento

El proceso de elaboración de un escalograma implica, más normalmente, los pasos que constan en la tabla 2 (García Llamas, 2001; Barbero, 1999; Buendía, 1998; Del Rincón y Otros, 1995; Arce, 1994; Sierra Bravo, 1992; Dotson y Summers, 1984).

TABLA 2
RESUMEN BÁSICO DE LOS PASOS PARA LA CONSTRUCCIÓN
DE UN ESCALOGRAMA GUTTMAN.

PASOS PARA DISEÑAR Y ANALIZAR UN ESCALOGRAMA GUTTMAN
1. Elaborar una batería de ítems procurando que sean «escalones cuasi-perfectos», al menos en teoría, apoyándose en alguna propuesta conceptual sobre el rasgo medido.
2. Ordenarlos estimando su dificultad ascendente (desde el primero —más fácil— hasta el último —el más difícil—); lógicamente es posible que haya escalones mal situados.
3. Seleccionar una muestra de población, lo más amplia posible y relacionada con los objetivos de medida del instrumento, a los que se les administra el escalograma.
4. Valorar la unidimensionalidad y consistencia interna de la escala.
5. Los datos obtenidos se codifican y ordenan en una matriz rectangular (1,0), con los sujetos en las filas y su valoración a los ítems en las columnas.
6. Para cada ítem, se recuenta el número de 1 que contiene. Con este valor se reordenan de forma estricta todos los ítems, desde el que obtiene el mayor valor (más fácil) hasta el que obtiene el menor valor (más difícil y, por tanto, con más 0).
7. Para cada sujeto se compara el patrón de respuesta observado con el patrón de respuesta ideal, y se cuentan los errores producidos en el conjunto de la matriz.
8. Comprobar si la escala que hemos construido es realmente una escala acumulativa para ello hay que realizar las siguientes comprobaciones:
8.1. Aplicar el Coeficiente de Reproductibilidad de Guttman (CR). Si el coeficiente de reproductividad es menor que 0.90, se considera que no hay un buen ajuste y habría que eliminar/re-elaborar ítems y comenzar de nuevo; seguir si es $\geq 0,90$.
8.2. Calcular la Reproductibilidad Marginal Mínima (RMM) para comprobar si el valor CR corresponde una escala bien realizada o es producto de puntuaciones marginales extremas debido a un sesgo poblacional o error de diseño.
8.3. Calcular el Coeficiente de Escalabilidad (CE) que debe resultar como mínimo 0.60 para indicarnos que la escala es verdaderamente acumulativa.
9. Elaborar el escalograma definitivo que usarán los sujetos para euto-posicionarse.

En nuestro caso, hemos considerado añadir un estudio específico de la validez concurrente y adecuación empírica del uso interpretativo que proponemos para este escalograma. Como señalan algunos autores (Jornet y Suarez, 1996; Hambleton, 1984), la validez de las medidas tiene mucho que ver con la finalidad y el uso que se va a hacer con ellas. Dado que el modelo de trabajo diagnóstico en la teleformación exige la aplicación por Internet del escalograma a los usuarios, siendo ellos los encargados de autoevaluarse y tomar la decisión sobre su nivel de dominio de las TID.

El estudio piloto, de carácter correlacional, se realiza sobre una muestra intencional final de 61 sujetos seleccionados por cuotas entre todos los cursos (1º, 2º, 3º, 4º, 5º y 3º Ciclo) que conforman la población de estudiantes de la Facultad de Ciencias de la Educación (Pedagogía) de la Universidad de Sevilla. En el proceso de elaboración del escalograma hemos partido de una propuesta empírica de 18 ítems (obviamente, muchos más de los necesarios), que se proponen en dos fases a los sujetos del estudio; primero, en una fase de revisión de la redacción de los ítems; y segundo, tras asegurar la comprensión básica de los mismos, en la fase de aplicación al conjunto de la muestra.

Una vez categorizados los datos (0,1) pudimos comprobar la inadecuación conceptual de algunos ítems en la escalera; pues, una vez ordenados según las respuestas de los sujetos, se comprobaba que algunos no eran requisito o «escalón» imprescindible para que los sujetos considerasen positivamente el siguiente. Una vez que se reduce conceptualmente la escala, aún es posible que algunos ítems sean impropios para una escala unidimensional (lo que se comprueba mediante análisis de componentes principales, así como valorando la correlación entre cada Ítem de la escala y el Total de la misma y calculando el coeficiente de consistencia interna Alpha de Cronbach). Finalmente, con el objeto de comprobar la validez concurrente se construye y valida como criterio una escala likert y se realizan análisis correlacionales y pruebas de contraste para identificar la adecuación del instrumento y de las normas de uso propuestas para el autodiagnóstico del nivel de dominio de las TID (según tres niveles: principiante, iniciado y experto). A continuación exponemos el proceso y resultados de tales procedimientos analíticos desarrollados sobre los datos con SPSS.

5. RESULTADOS ACERCA DE LA VALIDEZ DE CONSTRUCTO, FIABILIDAD Y REPRODUCTIBILIDAD DEL ESCALOGRAMA TID

Los resultados del estudio metodológico del escalograma se abordan en dos subapartados interrelacionados; de un lado, el cumplimiento de los criterios de unidimensionalidad y consistencia interna; de otro, el cumplimiento de los criterios empíricos de reproductibilidad y escalabilidad real del escalograma.

5.1. Unidimensionalidad y consistencia interna del escalograma

El análisis de la unidimensionalidad del escalograma procede mediante un análisis correlacional multivariante de componentes principales sobre los datos empíricos aportados por la muestra del estudio. La exploración tiene como objetivo hallar un componente principal que pueda explicar una cierta fracción de varianza conjunta

para todos los ítems del instrumento. Un resumen de los resultados obtenidos consta en las tablas 3 y 4 siguientes:

Como puede observarse en los estadísticos expuestos en la tabla anterior, puede decirse que la escala conformada por este conjunto de ítems es unidimensional; esto es, estamos midiendo niveles de un mismo rasgo con los diversos indicadores o ítems. Ello se comprueba porque, una vez que se rechaza la hipótesis de esfericidad (Bartlett $p \leq 0,001$) y que obtenemos una correcta evaluación de la medida de adecuación de muestreo (KMO = 0,71, que nunca debe ser inferior a 0,50), se extrae un único componente explicativo del 42,7% de la varianza en el cual se saturan adecuadamente el conjunto de los ítems del escalograma (saturaciones $\geq 0,30$). De forma complementaria, en la tabla 4 se aborda el estudio estadístico de la consistencia interna de la escala que componen este conjunto de ítems. Tiene especial interés la interpretación de la correlación del Ítem-total y el coeficiente Alpha de Cronbach.

TABLA 3
ANÁLISIS DE COMPONENTES PRINCIPALES DEL ESCALOGRAMA

ANÁLISIS DE COMPONENTES PRINCIPALES							
Medida de adecuación muestral de Kaiser-Meyer-Olkin						,712	
Prueba de esfericidad de Bartlett				Chi-cuadrado aproximado	127,040		
				grados de libertad		21	
				Nivel de Significación		,000	
Varianza total explicada							
Autovalores iniciales				Sumas de las saturaciones al cuadrado de la extracción			
Componente	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	
1	2,992	42,740	42,740	2,992	42,740	42,740	
2	,998	14,255	56,995				
3	,916	13,083	70,079				
4	,743	10,620	80,699				
5	,684	9,774	90,473				
6	,519	7,408	97,881				
7	,148	2,119	100,000				
Método de extracción: Análisis de Componentes Principales.							
MATRIZ DE COMPONENTES						COMPONENTE 1	
ITEM1* ¹						-----* ¹	
ITEM2						,463	
ITEM3						,453	
ITEM4						,800	
ITEM5						,672	
ITEM6						,866	
ITEM7						,627	
ITEM8						,582	
*El ítem 1, con varianza = 0, se excluye de los análisis por los motivos expuestos posteriormente cuando explicamos el cálculo de las categorías modales para el estudio del coeficiente de escalabilidad del escalograma.							

TABLA 4
CONSISTENCIA INTERNA DE LA ESCALA UNIDIMENSIONAL.

ESTUDIO DE LA CONSISTENCIA INTERNA DEL ESCALOGRAMA (Ítem - Total)					
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item- Total Correlation	Squared Multiple Correlation	Alpha if Item Deleted
ITEM2	1,5574	2,6842	,3312	,1352	,7539
ITEM3	1,7213	2,5710	,3231	,1360	,7636
ITEM4	2,0656	2,2956	,6188	,7033	,6873
ITEM5	2,0820	2,4765	,4822	,3736	,7207
ITEM6	2,0984	2,2235	,7282	,7497	,6615
ITEM7	2,1803	2,6503	,4689	,2268	,7247
ITEM8	2,2623	2,8967	,4297	,2518	,7380
Reliability Coefficients	Los 7 ítems (Ítem1 excluido de los análisis)* ¹				
Alpha =	,7534	Standardized item alpha =	,7625		

Se comprueba que el valor Alpha de Cronbach es superior a 0,75, límite aconsejable para este tipo de instrumentos, aunque puede ser más flexible en otros casos y situaciones de investigación (Morales, 2000).

5.2. Escalabilidad empírica real de la «escalera» que implica el escalograma

Una vez realizadas estas primeras manipulaciones se desarrolla un estudio específico mediante los pasos explicados anteriormente en la tabla 2. Para nuestra escala, éstos nos conducen a la construcción de los siguientes indicadores técnicos de la misma (el punto de partida es el recuento de errores sobre los datos empíricos obtenidos de la muestra para los ítems seleccionados, constan en la tabla 5):

Una vez recontados los 44 errores, se procede al cálculo del Coeficiente de Reproducibilidad. $CR = 1 - [n^{\circ} \text{ total de errores} / (n^{\circ} \text{ de ítems} * n^{\circ} \text{ de sujetos})] = 0,9098$.

Para comprobar si se produce una sobreestimación, procedemos al cálculo de las categorías modales (probabilidad empírica en cada ítem de la modalidad —0 ó 1- que más se repite). En este ejemplo; ítem1 = $?_{(0,57)}$, ítem2 = 0,77, ítem3 = 0,60, ítem4 = 0,73, ítem5 = 0,75, ítem6 = 0,77, ítem7 = 0,85 e ítem8 = 0,93; y, con ellas se desarrolla el coeficiente de Reproducibilidad Marginal Mínima (RMM = sumatorio de las categorías modales dividido por el número de ítems). Para el caso de nuestro estudio piloto, se aporta un valor empírico externo para la categoría modal del primer ítem o «escalón» (basado en el estudio diagnóstico de la muestra de alumnos —n=154- que cursan la asignatura). Ello, es un artificio metodológico que está bien justificado en nuestro estudio. Dado que nuestro objetivo es construir un escalograma general y válido para cualquier técnica e instrumento de diagnóstico, hemos forzado a los sujetos de la muestra para que valoraran algún instrumento conocido. Por ello, el escalón 1º (*¿Usted conoce este instrumento o estrategia de diagnóstico?*) pierde sentido y obtiene en todos los sujetos el valor 1 = SI, lo que supone que su categoría modal adquiere el valor probabilístico 1.00

TABLA 5
DATOS EMPÍRICOS DEL ESTUDIO PILOTO Y CÁLCULO DE ERRORES
(TOTAL=44)

Matriz de Datos del Estudio Piloto (61 casos)								
Nº	Datos	Errores	Nº	Datos	Errores	Nº	Datos	Errores
1	11111111	0	21	11010010	2	41	10100000	2
2	11111111	0	22	11100000	0	42	11000000	0
3	11111101	2	23	11100000	0	43	11000000	0
4	11111110	0	24	11100010	2	44	11000000	0
5	11111100	0	25	11100000	0	45	10100000	2
6	11111100	0	26	11100000	0	46	11000000	0
7	11101110	2	27	11100000	0	47	11000000	0
8	11110110	2	28	11100000	0	48	11000000	0
9	11110110	2	29	11100000	0	49	10100000	2
10	11110100	2	30	11100000	0	50	11000000	0
11	11101001	2	31	11100000	0	51	11000000	0
12	11011100	2	32	11100000	0	52	10100000	2
13	11110100	2	33	11010000	2	53	11000000	0
14	11110100	2	34	10101000	2	54	10000000	0
15	11011100	2	35	11100000	0	55	10000000	0
16	11101010	2	36	10101000	2	56	10000000	0
17	11101000	2	37	11100000	0	57	10000000	0
18	11110000	0	38	11001000	2	58	10000000	0
19	11100000	0	39	11000000	0	59	10000000	0
20	11100000	0	40	11000000	0	60	10000000	0
						61	10000000	0

—algo absurdo en realidad, pues *todos* los alumnos no tienen conocimiento de *todas* las técnicas de diagnóstico—.

Una vez realizada esta corrección, que nos conduce a considerar la categoría modal para el ítem1 = 0,57 correspondiente al estudio empírico en que se describe la autovaloración sobre el conocimiento general de los instrumentos de diagnóstico, se suma junto con las restantes categorías modales para calcular el RMM = $5,97 / 8 = 0,75$. Éste nuevo valor junto con el Coeficiente de Reproducibilidad de la escala permiten, finalmente, calcular el Coeficiente de Escalabilidad del escalograma ($CE = [(CR - RMM) / (1 - RMM)] = 0,64$) adecuado para considerar que los diversos ítems son pasos o escalones de una escala bien dimensionada y ordenada; esto es, que estamos verdaderamente ante un Escalograma Guttman.

6. RESULTADOS ACERCA DE LA VALIDEZ CONCURRENTE Y ADECUACIÓN EMPÍRICA DEL USO INTERPRETATIVO AUTORREGULADO PROPUESTO PARA EL ESCALOGRAMA TID

Hasta este punto, hemos mostrado el interés contextualizado del instrumento en el marco de una serie de objetivos diagnósticos implicados en los procesos de teleformación propuestos para innovar la materia «Técnicas e Instrumentos de Diagnostico» en el marco de la Sociedad del Conocimiento. También se han expuesto los estudios de fiabilidad y validez «de constructo» del escalograma junto con el proceso de elaboración del mismo. Igualmente, se ha presentado en un apartado anterior el instrumento definitivamente obtenido y «afinado» para su uso en el marco del sistema de teleformación S.T. MIDE, indicando los criterios de autoaplicación por los usuarios.

Sin embargo, el proceso de análisis empírico de dicho instrumento y de su adecuación no ha concluido todavía, principalmente, porque deben resolverse aún posibles dudas respecto de su utilidad real. Nos referimos, específicamente, a dos aspectos complementarios: 1) su adecuación y sensibilidad respecto de algún criterio concurrente de medida que nos ofrezca seguridad respecto de la validez de los «valores de rasgo» (nivel ordinal de dominio de las TID) producidos con este escalograma; y, 2) la adecuación empírica del uso interpretativo para tres categorías (principiante, iniciado y experto) propuesto a partir de los puntos de corte («grupos de escalones») conceptualmente justificados para su fácil autoaplicación por los usuarios del sistema.

Ambos aspectos remiten básicamente hacia dos aspectos de la investigación que se recogen en sendos subapartados siguientes: uno, sobre la construcción conceptual, fiabilidad y validez de una escala de tipo Likert que permite obtener medidas ordinales respecto del mismo rasgo estudiado sobre la misma muestra de sujetos; y otro, sobre la validez concurrente de las medidas realizadas con este nuevo instrumento y el escalograma, así como de la adecuación y discriminación empírica de los puntos de corte establecidos con vistas a la interpretación de los resultados que cada usuario obtiene con el escalograma y las instrucciones ofrecidas para su uso.

6.1. Construcción y validación de una escala Likert para el análisis del nivel de dominio de TID

Dado que el proceso de construcción de escalas tipo Likert está bastante extendido en la literatura científica —véase, por ejemplo, su exposición en Summers (1984), Sierra Bravo (1992), Arce (1994), Colás y Buendía (1998) y Morales (2000); entre otros— no nos detenemos en los aspectos básicos. La escala construida consta en la tabla 6, se aplica conjuntamente con el escalograma a la misma muestra del estudio piloto antes expuesta para cumplir su función como criterio de validez concurrente.

Respecto de la calidad de constructo, su unidimensionalidad, se demuestra empíricamente en los resultados del análisis factorial de componentes principales que se exponen a continuación. Como puede observarse en las tablas 7, 8 y 9, un solo factor agrupa al conjunto de los ítems de la misma (con saturaciones superiores a 0,60). Este factor explica un amplio porcentaje de la varianza (54%), dándose una óptima adecuación de muestreo ($KMO = 0,73$) y rechazándose la hipótesis de esfericidad.

TABLA 6
ESCALA LIKERT SOBRE DOMINIO DE LAS TID

RESPONDA SEGÚN SU GRADO DE ACUERDO (1= totalmente en desacuerdo / 5= totalmente de acuerdo)

USTED CONOCE SUFICIENTEMENTE LA TEORÍA DE ESTA TÉCNICA	1	2	3	4	5
USTED CONOCE SUFICIENTEMENTE SU USO PRÁCTICO EN EDUC.	1	2	3	4	5
USTED CONOCE SUFICIENTEMENTE SU ELABORACIÓN TÉCNICA	1	2	3	4	5
USTED MANEJA DICHA TÉCNICA CON VERSATILIDAD PRÁCTICA	1	2	3	4	5
USTED PUEDE FORMAR A OTROS PROFESIONALES EN SU USO	1	2	3	4	5
USTED CONOCE LÍNEAS DE INVESTIGACIÓN SOBRE ESTA TÉCNIC.	1	2	3	4	5
USTED SABE INVESTIGAR SOBRE DICHA TÉCNICA EN EDUCAC.	1	2	3	4	5

TABLA 7
KMO Y PRUEBA DE BARTLETT PARA LOS DATOS DE LOS ÍTEMS DE LA ESCALA LIKERT

KMO y prueba de Bartlett		
Medida de adecuación muestral de Kaiser-Meyer-Olkin.		,730
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	185,274
	gl	21
	Sig.	,000

TABLA 8
PRUEBA DE UNIDIMENSIONALIDAD Y VARIANZA TOTAL EXPLICADA POR EL COMPONENTE PRINCIPAL

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	3,787	54,095	54,095	3,787	54,095	54,095
2	,979	13,988	68,083			
3	,869	12,409	80,492			
4	,562	8,022	88,514			
5	,373	5,333	93,847			
6	,259	3,707	97,554			
7	,171	2,446	100,000			

Método de extracción: Análisis de Componentes principales.

TABLA 9
GRADO DE SATURACIÓN DE LOS ÍTEMS SOBRE EL COMPONENTE PRINCIPAL

Matriz de componentes

	Componente 1
ITEM1LIKERT	,658
ITEM2LIKERT	,756
ITEM3LIKERT	,821
ITEM4LIKERT	,740
ITEM5LIKERT	,821
ITEM6LIKERT	,648
ITEM7LIKERT	,683

La fiabilidad de la medida obtenida con esta escala Likert, interpretada desde la consistencia interna de sus elementos y valorada mediante el coeficiente Alpha de Cronbach, consta en la tabla 10 siguiente. Sobre esta puede observarse la alta fiabilidad obtenida en la prueba piloto del instrumento ($\text{Alpha} = 0,85$), también las altas correlaciones Ítems-Total obtenidas contribuyen a dicha interpretación y a la valoración del mantenimiento de la escala en su versión completa. Ello supone que se puede proceder al calculo de la nueva variable o distribución de datos que se produce por la suma de los valores de posicionamiento ordinal de los sujetos de la muestra sobre los 7 ítems de la escala Likert, el valor obtenido es un indicador ordinal de su nivel de dominio de las TID según esta escala (7-35). Esta nueva variable medida constituye el criterio buscado para analizar la validez concurrente, siendo evidente la validez de contenido si se comparan e interpretan los distintos ítems a la luz de los objetivos del diagnóstico. Las medidas ordinales obtenidas para cada sujeto del estudio piloto en esta escala sumativa se utilizan a continuación como criterio de validez concurrente respecto de las obtenidas con el escalograma Guttman.

La distribución de valores producidos por la escala Likert se describe en el gráfico 1 subsiguiente, como paso previo a su estudio correlacional respecto de los valores producidos por el escalograma (gráfico 2).

Puesto que vamos a considerar esta distribución de valores como criterio en el desarrollo de diversas pruebas conviene plantear el posible uso de técnicas paramétricas con base en estos datos. Es evidente que violamos el supuesto paramétrico del nivel de medida (que debe ser como mínimo de intervalo), pero queremos saber si podemos simular dichos supuestos desde la base de los dos restantes criterios. El primero se cumple, pues estamos ante una muestra grande (57 casos válidos) con un

TABLA 10
ESTUDIO DE LA FIABILIDAD (CONSISTENCIA INTERNA MEDIANTE ALPHA DE CRONBACH) DE LA ESCALA LIKERT.

Item-total Statistics					
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item- Total Correlation	Squared Multiple Correlation	Alpha if Item Deleted
LIKERT1	14,2456	28,9743	,5300	,5084	,8492
LIKERT2	13,7544	26,9743	,6498	,6187	,8333
LIKERT3	13,9649	25,3559	,7282	,5937	,8209
LIKERT4	14,3509	25,8390	,6233	,6260	,8375
LIKERT5	14,8070	24,8728	,7305	,6677	,8202
LIKERT6	14,8596	28,9442	,5338	,4701	,8488
LIKERT7	14,4386	27,0006	,5658	,4746	,8455
R E L I A B I L I T Y A N A L Y S I S - S C A L E (A L P H A)					
Reliability Coefficients 7 items					
Alpha = ,8570 Standardized item alpha = ,8565					

rango de valores posibles amplio ($35-7 = 28$). El segundo se refiere a la probabilidad de que tales valores se distribuyan de forma aproximadamente normal. Este criterio también se cumple, como demuestra la prueba de Kolmogorov-Smirnov ($p \geq 0.05$, pues no se desea rechazar la hipótesis nula de la diferencia de distribuciones) que expone la tabla 11.

Una vez que conocemos las propiedades de la distribución de datos producidos con la escala Likert y que esta aporta indicadores fiables y válidos en la medida realizada, contamos con un criterio para estudiar la validez concurrente del escalograma. El apartado subsiguiente aborda dicha faceta de este estudio empírico.

6.2. Validez concurrente del escalograma y sus instrucciones de uso e interpretación en el marco del S.T. MIDE

El proceso de valorar la validez concurrente del escalograma implica considerar la correlación existente entre los valores de medida producidos por medio de la escala Likert y los elaborados mediante el propio escalograma. Si ambos miden el mismo rasgo latente o constructo considerado teóricamente (nivel ordinal de dominio de las

TABLA 11
 PRUEBA DE KOLMOGOROV-SMIRNOV (NORMAL) PARA LA MUESTRA DE VALORES PRODUCIDOS CON LA ESCALA LIKERT

N (casos válidos para este análisis)		57
Parámetros normales(a,b)	Media	16.7368
	Desviación típica	5.9747
Diferencias más extremas	Absoluta	.114
	Positiva	.114
	Negativa	-.073
Z de Kolmogorov-Smirnov		.861
Sig. asintót. (bilateral)		.449
a La distribución de contraste es la Normal; b Se han calculado a partir de los datos.		

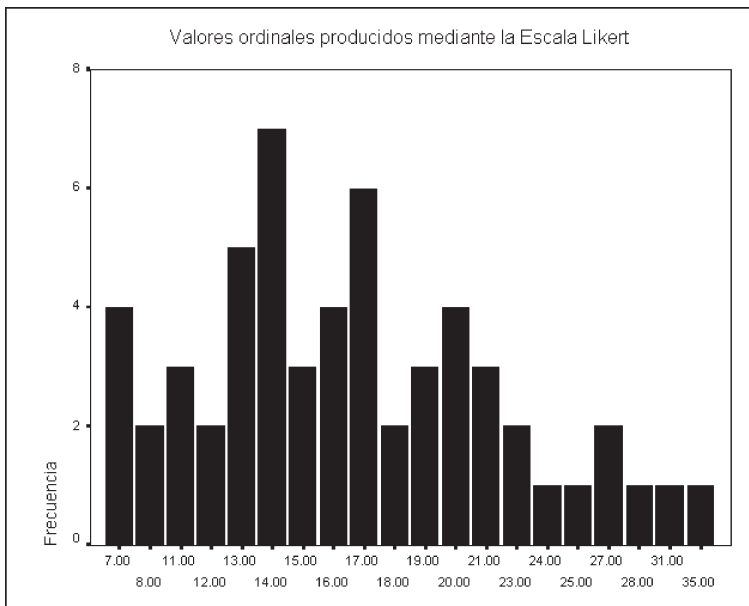


Gráfico 1

Distribución de los valores producidos mediante la escala Likert.

TID), sus distribuciones deben mostrar altas correlaciones positivas. A continuación, en la gráfica 2 presentamos la distribución de datos producidos con el escalograma. Dicha distribución de datos va a ser truncada a posteriori, con el objeto de valorar los niveles correspondientes a las categorías conceptuales elaboradas con base al esca-

lograma. Ello supone considerar el agrupamiento de valores según el criterio: ($\leq 2 = 1 =$ Principiante; > 2 y $\leq 5 = 2 =$ Iniciado; $y, \geq 6 = 3 =$ Experto). Dicho agrupamiento deviene del análisis teórico de los «escalones» del escalograma para aportar una vertiente interpretativa sencilla y didácticamente útil a los usuarios del sistema de teleformación.

No obstante, antes de dicha simplificación podemos considerar la validez de la medida (en la escala 1-8) según su concurrencia con los datos aportados por la escala Likert. En este sentido, las altas correlaciones obtenidas confirman la validez concurrente de ambas distribuciones de valores del rasgo medido. Concretamente, se obtienen altos valores para los coeficientes r de Pearson (0,761) y rho de Spearman (0,711), siendo ambos significativos a un nivel de significación $p \leq 0,001$. Precisamente, el gráfico 3 siguiente recoge el diagrama de dispersión que cruza ambas variables.

En el diagrama de dispersión se puede observar claramente la relación lineal positiva de ambas variables según la distribución de la nube de puntos. Asimismo, en el gráfico 4 se recogen los diagramas de caja y bigotes de los grupos conformados (principiantes, iniciados y expertos) con el escalograma una vez truncada la distribución de la variable original. En este diagrama se describen los valores obtenidos por cada grupo en la escala Likert, poniendo nuevamente ambos instrumentos en relación lineal.

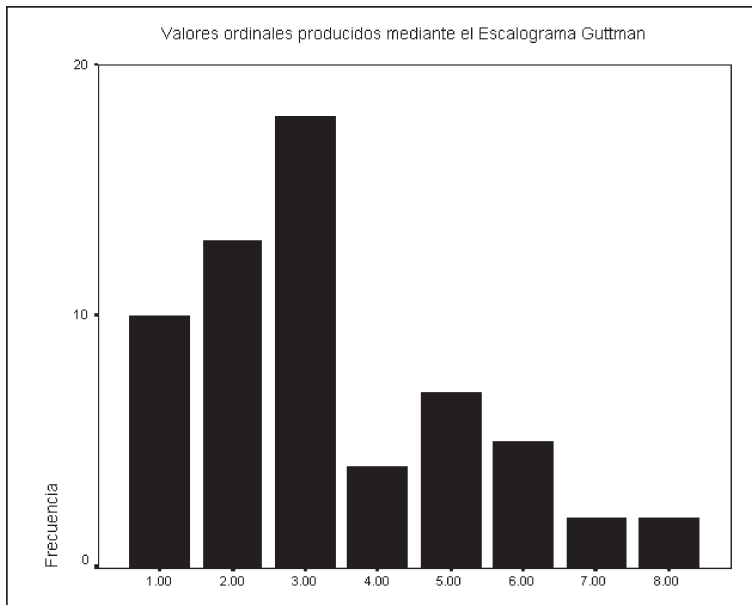


Gráfico 2
Distribución de valores producidos por el escalograma Guttman.

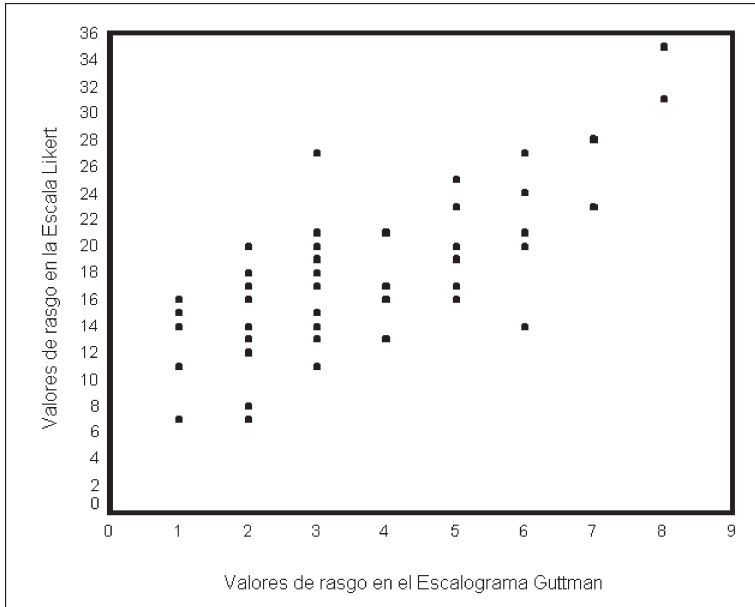


Gráfico 3

Diagrama de dispersión entre los datos producidos por ambos instrumentos de medida.

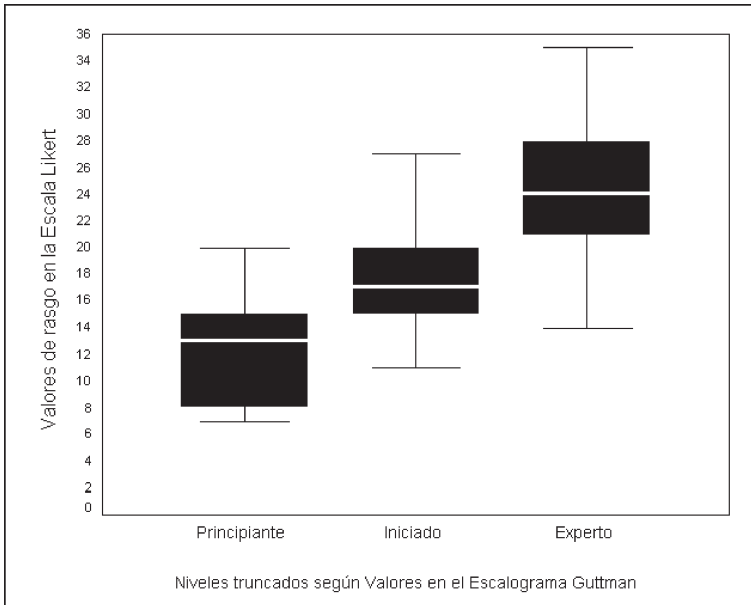


Gráfico 4

Diagramas de caja y bigotes para cada nivel categorial del escalograma Guttman.

Dicho diagrama de caja y bigotes expresa mediante un gráfico de resumen expresivo de cada grupo: la mediana, los cuartiles y los valores extremos respecto de la variable usada como criterio. La caja representa la amplitud intercuartil que contiene el 50% de los valores centrales. Los «bigotes» son las líneas que se extienden desde la caja hasta los valores más altos y más bajos, excluyendo los valores atípicos. Una línea blanca que atraviesa cada una de las cajas indica el valor de la mediana de cada grupo. Como puede observarse en la tabla 12, en cada grupo obtenemos medias expresivamente distintas, que caracterizan los niveles de dominio expresado ordinalmente con cada instrumento de diagnóstico.

Como expone la tabla 13 el análisis de la varianza (ANOVA) dichas diferencias de medias son significativas ($p \leq 0,001$), tanto globalmente como en la comparación múltiple de cada grupo con los restantes mediante la prueba de Scheffé (tabla 14).

TABLA 12
 ESTADÍSTICOS DESCRIPTIVOS PARA LOS GRUPOS: 1=PRINCIPIANTE; 2=INICIADO; Y, 3=EXPERTO EN EL ANOVA

	N	Media	Desviación típica	Error típico	Intervalo de confianza para la media al 95%		Mínimo	Máximo
					Límite inferior	Límite superior		
1.00	22	12.4545	3.8635	.8237	10.7416	14.1675	7.00	20.00
2.00	26	17.5769	3.8280	.7507	16.0307	19.1231	11.00	27.00
3.00	9	24.7778	6.2805	2.0935	19.9502	29.6054	14.00	35.00
Total	57	16.7368	5.9747	.7914	15.1515	18.3222	7.00	35.00

TABLA 13
 ANOVA EN LA ESCALA LIKERT PARA LOS TRES GRUPOS CONFORMADOS MEDIANTE EL ESCALOGRAMA

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	1003.696	2	501.848	27.226	.000
Intra-grupos	995.356	54	18.433		
Total	1999.053	56			

TABLA 14
COMPARACIONES MÚLTIPLES (SCHEFFÉ) ENTRE LOS GRUPOS
DE PRINCIPIANTES, INICIADOS Y EXPERTOS

		Diferencia de medias (I-J)	Error típico	Sig.	Intervalo de confianza al 95%	
					Límite inferior	Límite superior
1.00	2.00	-5.1224(*)	1.2437	.001	-8.2531	-1.9917
	3.00	-12.3232(*)	1.6988	.000	-16.5995	-8.0470
2.00	1.00	5.1224(*)	1.2437	.001	1.9917	8.2531
	3.00	-7.2009(*)	1.6604	.000	-11.3805	-3.0212
3.00	1.00	12.3232(*)	1.6988	.000	8.0470	16.5995
	2.00	7.2009(*)	1.6604	.000	3.0212	11.3805

* La diferencia entre las medias es significativa al nivel .05

En resumen, puede interpretarse que el escalograma y las instrucciones de uso que proponemos para el mismo, cara a la función de autoevaluación diagnóstica del nivel (principiante, iniciado y experto) de dominio de las «Técnicas e Instrumentos de Diagnóstico» (TID), son válidas y fiables, conceptualmente adecuadas y sensibles. Produciendo resultados de medida acordes con el nivel de rasgo que deseamos medir en los sujetos. El nivel de dominio autoevaluado queda convenientemente diferenciado y escalonado en los tres niveles propuestos para el diseño del sistema de teleformación.

Estas valoraciones desarrolladas en tres niveles ordinales (1-3); por tanto más globales, pese a que son una simplificación de los resultados obtenidos por el escalograma, mantienen los altos niveles de discriminación entre sus categorías y altas correlaciones con el criterio. Al igual que la versión inicial del escalograma (1-8), se producen valores altamente correlacionados con las medidas elaboradas como criterio mediante la escala tipo Likert ($r = 0,704$ y $\rho = 0,680$; siendo ambos coeficientes significativos a un nivel de $p \leq 0,001$ para el agrupamiento en los tres niveles descritos).

7. CONCLUSIONES

Como conclusión de este estudio empírico y metodológico sobre el escalograma Guttman construido, podemos resaltar tres bloques de comentarios: El primero *sobre el instrumento que ya disponemos* para su uso en la teleformación de estudiantes universitarios que cursan las enseñanzas de la asignatura «técnicas e instrumentos de diagnóstico», así como para otros posibles usuarios que accedan a dicho sistema con el objeto de formarse en estos contenidos metodológicos. Este consta en las páginas que el Laboratorio LIC (MIDE - Universidad de Sevilla) cuenta en las páginas institucionales de nuestro campus virtual: <http://www.us.es/lablic/STMIDE/STMIDE.html>. El segundo bloque de comentarios se refiere a los resultados finales obtenidos con el estudio expuesto y que conducen a la exposición final de un conjunto de *característi-*

cas técnico-metodológicas del instrumento. Finalmente, el tercer bloque de comentarios lo reservamos para desarrollar un conjunto de aspectos que consideramos propuestas oportunas para debatir respecto de la «idea de fondo» que conduce este trabajo y que no es otra que *potenciar el discurso sobre el impacto y los cambios metodológicos auspiciados por el nuevo contexto educativo* que implica el advenimiento de la Sociedad del Conocimiento; éstos que tanto a la didáctica como al diagnóstico.

Respecto de la primera línea de conclusiones, debemos resaltar que ya contamos con el instrumento de diagnóstico que necesita el S.T. MIDE. Este tiene por objetivo la evaluación diagnóstica inicial autorregulada por los usuarios del sistema de teleformación S.T. MIDE. Esto se plantea con el objetivo de determinar, de una forma rápida y sencilla, el «nivel de acceso» a los contenidos y actividades de los usuarios de dicho sistema (obviamente, para ello el sistema debe estar diseñado para diversos niveles de usuarios). El escalograma sobre TID estudiado en este artículo cumple con dicho objetivo, engrosando el conjunto de técnicas e instrumentos disponibles para el trabajo virtual de la materia.

Hay que tener en cuenta que todo sistema de teleformación tiene la misión de «regular los aprendizajes», una tarea que en la enseñanza presencial asume el profesor. Esta regulación «virtual» del usuario es lo que diferencia a un verdadero sistema de (tele)formación de un simple sistema de (tele)información. En este tipo de enseñanza el profesor está presente a través de las páginas Web que diseña y que contienen los mecanismos de regulación educativa necesarios. En este caso, se trata de que cuando un sujeto «entra» al sistema de teleformación, este le ofrece un mecanismo para autoconocerse un poco mejor y orientar las estrategias de aprendizaje que puede desarrollar con mayor éxito y propiedad en función de sus necesidades. Esta evaluación diagnóstica inicial no debe ser compleja y agotar la inversión de esfuerzo de los usuarios, además tiene como objetivo sencillamente «orientar» el acceso al nivel didáctico adecuado, más que «controlar» el nivel formativo para su certificación o promoción. Todo ello justifica una autoevaluación mediante un instrumento como el escalograma, que es rápido, cómodo y fácil de interpretar según las instrucciones de uso.

En nuestro caso, el escalograma plantea tres niveles de diagnóstico (y por tanto de acceso a la metodología didáctica del S.T. MIDE): 1) Nivel de Principiante, vinculado fundamentalmente al reconocimiento del amplio abanico de técnicas e instrumentos de diagnóstico actualmente disponible y a su capacidad de aplicación básica en una situación concreta. 2) Nivel de Iniciado, vinculado a la capacidad de elaborar, seleccionar, aplicar y estudiar metodológicamente las características fundamentales de cada tipo de técnica e instrumento; reconociendo los fundamentos básicos y las concepciones teóricas que los sostienen. 3) Nivel de Experto, que ya implica un amplio conocimiento crítico de la materia y la capacidad de investigar sobre la misma desarrollando nuevas teorías, técnicas e instrumentos; así como, la capacidad para formar a técnicos y profesionales del campo educativo que deben aplicarlas. El S.T. MIDE incluye 432 actividades formativas diseñadas para esta diversidad de usuarios.

El escalograma sobre TID construido y descrito en este artículo tiene como rasgo fundamental su facilidad de utilización por los usuarios del sistema, sean cuales sean sus niveles formativos previos; pues sólo se deben preocupar de responder con sinceridad (que está asegurada, pues nadie les controla ni tiene sentido elegir actividades didácticas impropias, más fáciles o más difíciles, respecto del propio nivel de desarro-

llo formativo) e identificar en que «escalón» tienen problemas educativos no resueltos, siendo ese su nivel formativo. Dicho escalón se identifica inmediatamente con una de las tres categorías descritas (principiante, iniciado y experto). El propio sistema guía la selección de actividades para su nivel de dominio de las TID autodiagnosticado.

Abordando el segundo bloque de conclusiones sobre el estudio empírico y metodológico del escalograma Guttman construido sobre las TID, podemos establecer las siguientes características técnicas del mismo. Para ello y partiendo de lo explicado en los apartados de resultados, recopilamos de forma muy resumida los resultados obtenidos del proceso de elaboración y validación empírica del mismo.

El escalograma Guttman sobre TID:

1. Es una escala unidimensional (1 CP, 42,7% varianza; altas saturaciones).
2. Los elementos (ítems) que lo integran presentan un elevado coeficiente de consistencia interna (Alpha de Cronbach > 0,75).
3. El Coeficiente de Reproducibilidad (CR = 0,91) de la escala es adecuado.
4. No se observa sobreestimación del valor de la escala, hallándose un apropiado Coeficiente de Escalabilidad (CE = 0,64).

Además, se ha construido otro instrumento (escala Likert) con el objeto de usarlo como criterio de validez concurrente en la medida del mismo rasgo diagnóstico. Este también presenta adecuadas características técnicas:

1. Unidimensionalidad (1 CP, 54% varianza; altas saturaciones).
2. Elevado coeficiente de consistencia interna (Alpha de Cronbach = 0,85).
3. Sus datos se distribuyen de forma aproximadamente normal ($p_{[K-S]} = 0.449$).

Las correlaciones observadas sobre los valores de medida de ambos instrumentos son suficientemente altas ($r_{\text{Pearson}} = 0,76$; $\text{Rho}_{\text{Spearman}} = 0,71$) y significativas ($p \leq 0,001$). Asimismo, respecto de las instrucciones de uso del escalograma (truncado para los niveles de «principiante», «iniciado» y «experto») se observan escalones muy bien diferenciados en relación con las puntuaciones de la escala Likert utilizada como criterio concurrente. Ello se muestra gráficamente en el diagrama de caja y bigotes; e igualmente, se contrasta estadísticamente mediante la prueba de ANOVA global ($p_{\text{Snedecor}} \leq 0,001$) como en la comparación múltiple de todos los subgrupos mediante la prueba de Scheffé (en todas las comparaciones $p \leq 0,001$). También, es muy expresiva de la validez de las instrucciones de autoaplicación previstas, la escasa pérdida de correlación de esta simplificación de los datos producidos por el escalograma en los tres subgrupos respecto de los datos de la escala Likert que ejerce como criterio de validez concurrente altas ($r_{\text{Pearson}} = 0,704$; $\text{Rho}_{\text{Spearman}} = 0,680$), también con un nivel de significación adecuado ($p \leq 0,001$). En resumen, para este segundo apartado de conclusiones, podemos decir que el escalograma se muestra adecuado y sensible para la función diagnóstica prevista, además se demuestra tanto fiabilidad como validez de constructo y concurrente al obtener valores de medida.

Finalmente, a modo de tercer bloque de conclusiones para el debate, proponemos las escalas de tipo Guttman (escalogramas) como medio idóneo para la evaluación

diagnóstica de los usuarios de los sistemas de teleformación. Estos son de aplicación rápida y cómoda, siendo muy orientativos en la regulación del proceso de selección de parrillas didácticas en el S.T. MIDE a la medida de cada usuario. Son estas propiedades las que justifican la aplicación de los escalogramas como instrumento de diagnóstico inicial en los procesos de teleformación. Aunque son algo difíciles de elaborar, en la actualidad dicha característica pierde importancia; en tanto se consideran más importantes sus propiedades de uso que sus indudables dificultades de diseño. Una de las fuentes en castellano más reconocibles y compartidas sobre la medición de actitudes es el texto, ya reeditado, de Pedro Morales Vallejo (2000); en éste, se recoge la sensibilidad positiva, que compartimos, hacia las escalas Likert. Pero se repasan muy someramente otras técnicas como los diferenciales semánticos, la metodología Q o las escalas acumulativas tipo Guttman, entre otras. De éstas, se señala que *«son de confección laboriosa y útiles para medir objetivos muy limitados. No es frecuente encontrar escalas de este tipo; a juicio de Kerlinger (1975, p. 350), son menos útiles y tienen menor aplicabilidad»* (Morales, 2000: 46). A partir de esta opinión ni siquiera se entra en la comparación con otras como las técnicas de Thurstone o Likert. Sin embargo, en el campo educativo, el texto de Colás y Buendía (1998) la recoge más extensamente e incluso con ejemplos y proceso de construcción (Buendía, 1998: 214-225). Igualmente, el texto de Del rincón y Otros (1995) recoge con precisión el proceso de elaboración, aunque cita la opinión negativa de Kerlinger antes expuesta.

Nuestra posición, después de un arduo trabajo de rediseño de contenidos de teleformación y adaptación de los procesos educativos y su regulación a la mediación de Internet, es que debemos *repensar y debatir algunas de las ideas establecidas sobre las técnicas de investigación, diagnóstico y evaluación* en general, y más particularmente, respecto de las escalas Guttman y su funcionalidad en el contexto educativo de la Sociedad del Conocimiento. Más recientemente, el texto de García Llamas y Otros (2001) también recoge este instrumento diagnóstico con cierta extensión, además estos autores señalan la utilidad de esta técnica para la simplificación de otras escalas, una función idónea actualmente que no valen ya, tanto, las pesadas evaluaciones del pasado.

Precisamente, el modelo didáctico tecnológico actual (especialmente en la enseñanza universitaria) se caracteriza por ofrecer mayor importancia a las actividades de los aprendices que a la actividad de los formadores, que son auxiliares de sus procesos de aprendizaje. Especialmente ahora, que este cambio de enfoque (centrado en el aprendizaje) nos llega bajo la tecnología del «Espacio Europeo de Educación Superior y el Crédito Europeo». Por ello, se justifica el cambio de «mentalidad» y el abandono de la idea de la «dificultad de elaboración de los escalogramas» como criterio de su desuso actual. También se justifica, especialmente, la valoración de posibles cambios en todos los procesos de evaluación diagnóstica y del papel que en ello tienen los propios usuarios/alumnos para conseguir que, progresivamente, los sistemas de teleformación respondan a las necesidades reales y diversas de sus usuarios. Éstos son hoy múltiples (quizás millones de posibles usuarios) y geográfico-culturalmente dispersos y que requieren de la atención (distante y mediada por NTICs) de nuestras universidades.

Como conclusión final y más concreta, respecto de nuestro proyecto de innovación docente basado en nuevas tecnologías, podemos decir que hoy contamos con un nuevo entorno de contenidos en Internet para el desarrollo de tareas combinadas (blen-

ded-learning) de enseñanza-aprendizaje presencial y virtual. Ello sitúa, al menos así lo creemos, nuestro trabajo en la sociedad del conocimiento y el tercer entorno de comunicación y aprendizaje. Finalmente, a partir de este estudio concreto, contamos expresamente con un escalograma válido y fiable para los objetivos de evaluación diagnóstica previstos en la materia «Técnicas e Instrumentos de Diagnóstico (TID)».

8. REFERENCIAS BIBLIOGRÁFICAS

- Arce, C. (1994): *Técnicas de construcción de escalas psicológicas*. Madrid: Síntesis Psicología.
- Ander-Egg, E. (1996): *Técnicas de investigación social*. México: El Ateneo.
- Barbero, M. I. (1999): *Psicometría II. Métodos de elaboración de escalas*. Madrid: UNED.
- Buendía, L. (1998): Técnicas e instrumentos de recogida de datos. En P. Colás y L. Buendía. *Investigación Educativa*. Sevilla: Alfar. 201-248.
- Castaño, C. y román, C. (2002): *Andalucía ante la sociedad de la información*. Sevilla: Consejo Económico y Social de Andalucía (CES). Junta de Andalucía.
- Colás, P. (2002). La investigación educativa en la (nueva) cultura de la sociedad del conocimiento. *XXI Revista de Educación*, 4, 77-93.
- Colas, P. y buendía, L. (1998): *Investigación educativa*. Sevilla. Alfar.
- Cornella, A. (2002): La tensión aprender, desaprender: crear valor. En *II Congreso de Aplicación de las Nuevas Tecnologías en la Docencia Presencial y E-Learning*. Valencia. Universidad Cardenal Herrera-CEU. (www.uch.ceu.es).
- Del Rincón y Otros (1995): *Técnicas de investigación en Ciencias Sociales*. Madrid: Dykinson.
- Dotson, L.E. y Summers, F. (1984): Cómo elaborar escalas técnicas de Guttman. En Summers, F. (Comp.) *Medición de actitudes*. México: Trillo.
- García Llamas, J. L. y Otros (2001): *Introducción a la investigación en educación (tomo II)*. Madrid: UNED.
- García Pérez, R. y Otros (2001): Diseño y evaluación de un programa de Informática Aplicada a la Investigación Educativa. *Fuentes*, 3, 193-218.
- García Pérez, R. (2002): Sistemas de Teleformación en la Enseñanza Universitaria Presencial: Experimentación de un modelo didáctico. *QuadernsDigitals*, 28. (www.quadernsdigitals.net).
- González García y Otros (1996): *Ciencia, Tecnología y Sociedad*. Madrid: Tecnos.
- Guttman, L. (1950): The basis for Scalogram analysis. En S. Stanffer y Otros. *Measurement and Prediction*. New Jersey: Princeton University Press.
- Guttman, L. (1984): Base para elaborar escalas con datos cualitativos. En F. Summers (comp.): *Medición de actitudes*. México: Trillo.
- Hambleton, R. K. (1984): Validating the tests scores. En R. BERK; *A guide to Criterion-Referenced Tests construction*. Baltimore: J. H. University Press.
- Jornet, J.M. y Suarez, J.M. (1996): Pruebas estandarizadas y evaluación del rendimiento: usos y características métricas. *Revista de Investigación educativa*, 14(2), 141-163.
- Morales, P. (2000): *Medición de actitudes en psicología y educación: construcción de escalas y problemas metodológicos*. Madrid. Universidad Pontificia Comillas.
- Sierra Bravo, R.(1992): *Técnicas de investigación social: teoría y práctica*. Madrid: Paraninfo.
- Summers, F. (1984): *Medición de actitudes*. México: Trillo.