Jornet, J.M y Suárez, J.M. (1989): Conceptualización del dominio educativo desde una perspectiva integradora en Evaluación Referida al Criterio (ERC). Bordón, v. 41.

CONCEPTUALIZACION DEL DOMINIO EDUCATIVO DESDE UNA PERSPECTIVA INTEGRADORA EN EVALUACION REFERIDA AL CRITERIO (ERC)

J. M. JORNET y J. M. SUÁREZ [*]

0. INTRODUCCIÓN

El tema de la definición del dominio es quizá, junto con el de la determinación de estándares, el tema central sobre el que gira la construcción de un Test Referido al Criterio (TRC). Así, este tema probablemente sea uno de los pocos en los que hay un gran acuerdo respecto a su trascendencia. Desde la definición de Glaser (1963) o el trabajo de Hively *et al.* (1968), pasando por los trabajos clásicos de Popham y Huseck (1969), Popham (1974, 1975, 1978), Millman (1974), hasta Nitko (1980, 1984) o Hambleton (1984, 1986), todos coinciden en afirmar que es la calidad de la definición del dominio la que va a posibilitar referir las puntuaciones individuales a criterios internos a la tarea o, lo que es lo mismo, a criterios de «calidad» definidos como de mínima competencia. Asimismo, debe entenderse que este tema es crucial, dado que se imbrica dentro del problema de la definición de la validad de contenido y en la de constructo, así como en otros conceptos de validez que han sido sugeridos para los TRC, como con curricular e instruccional (De la Orden, 1983; Madaus, 1983; McClung, 1978; Schmidt *et al.*, 1983; Linn, 1983, y Yalow y Popham, 1983).

En esta línea, se ha observado en las décadas de los sesenta y setenta la aportación de un gran número de referencias y recomendaciones en relación a los medios técnicos y de procedimiento que podían aumentar la calidad en la definición del dominio. Sin embargo, la gran mayoría de

[*] JESÚS M. JORNET y JESÚS M. SUÁREZ. Profesores titulares en el Departamento de Métodos de Investigación y Diagnóstico en Educación. Facultad de Filosofía y Ciencias de la Educación de la Universidad de Valencia.

trabajos dentro de los TRC giraban en torno a la dirección marcada por Hively y colaboradores, o bien con otras modificaciones inspiradas en acercamientos basados en objetivos educativos, uno de cuyos exponentes de mayor trascendencia son los objetivos amplificados de Popham. Sin embargo, ha sido ya en esta década, a nuestro entender, en la que se ha presentado un atisbo de sistematización respecto a aproximaciones útiles para la definición y estructuración del dominio con el trabajo de Roid y Haladyna (1982).

Por otra parte, en relación con los métodos para analizar la estructura del dominio, sin embargo, se ha trabajado muy poco propiamente en el ámbito de los TRC. Las recomendaciones al respecto han sido sugeridas normalmente desde otros ámbitos y exponente de ello es que tan sólo Millman (1974, 1980, 1984) o Hambleton (1984) han comentado este problema.

En este artículo vamos a revisar, en primer lugar, una conceptualización del dominio como universo de medida; a partir de ésta, se analizan los elementos implicados tanto en la definición como en la estructuración del dominio. Finalmente, efectuamos una pequeña síntesis de los medios que han sido utilizados en el marco de la Evaluación Referida al Criterio (ERC) para hacer operativo el acceso y manipulación de dominios, es decir, los sistemas que han sido aplicados para gestionar la enorme cantidad de información que supone un dominio definido, especificado y estructurado.

1. Algunas notas de reflexión conceptual

En un trabajo anterior (Jornet, 1987) intentamos plantear, desde un punto de vista teórico, la dicotomía existente entre una definición del dominio educativo desde el punto de vista del «educador» y desde las posiciones métricas. Es obvio que la dificultad implícita en una conceptualización genérica del dominio cercana a la realidad educativa plantea graves problemas a la medición, sobre todo cuando lo que se pretende es determinar un nivel mínimo de competencia. De este modo, en la ERC algunos autores, desde el ámbito de la determinación de estándares [1], han promovido asunciones restrictivas a la conceptualización del dominio, como es el caso de la homogeneidad, lo cual les ha llevado a plantear dominios de un solo objetivo y mediciones de un solo *item* (o, a lo sumo, dos o tres). De este modo, nosotros pensamos que si la realidad educativa es compleja y multidimensional, no pueden asumirse modelos de medición que restrinjan la consideración de la misma a situaciones exclusivamente parcializadas. Por ello, es mejor asumir el problema de definir estrategias, métodos y procedimientos de acercamiento a la medición educativa, antes que obviarlos con asunciones restrictivas. En consecuencia, en este apartado —y a lo largo del presente artículo— pretendemos aportar una visión del dominio educativo genérica, que sirva para acercarse

[1] Por ejemplo, autores que sustentan modelos de estado (Emrick y Adams, 1969; Emrick, 1971; Bart, 1973, 1975; Roudabush, 1974; Macready y Dayton, 1977, 1980 *a* y *b*), o bien continuos, basados en la teoría de la decisión (Lewis *et al.*, 1973; Huynh, 1976, 1980; Novick y Jackson, 1974; Swaminathan *et al.*, 1975; Van der Linden, 1980, 1984; Mellenbergh, Koppelaar y Van der Linden, 1977; Van der Linden y Mellenbergh, 1977; Novick y Lindley, 1978; Mellenbergh y Van der Linden, 1981).

al problema de la medición desde una óptica más consecuente con la realidad educativa. Así, pasamos a definir algunas notas de reflexión respecto a la conceptualización del dominio educativo como universo de medida.

Es evidente que por dominio se entiende el conjunto de todos aquellos elementos (objetivos, acciones, tareas, *items*) que representan el propósito de la instrucción. Esta consideración es genérica y, como tal, es aplicable a cualquier tipo de acercamiento (instrucción, metodología educativa, experimental, métrico, etc.) [2].

Ahora bien, desde un punto de vista métrico, el dominio constituye, por sí mismo, el universo de medida, el cual incluye todas las unidades que exhaustivamente lo definen, y éstas, a su vez, presentan una estructura bien determinada. A este respecto, consideramos que en el dominio pueden identificarse dos niveles en su conceptualización: *a)* definición y *b)* estructuración. Pasamos a comentar brevemente estos tópicos.

1.1. *Definición del dominio*

Un dominio está definido si están especificadas sus unidades [3]. Como unidades podemos clasificar: objetivos e *items*. De esta forma, se puede entender que la calidad de la definición del dominio está en relación con la concreción de las unidades que lo definen. Es obvio que una definición por objetivos es ambigua, mientras que los *items* representan las unidades mínimas de definición. En este sentido, la definición del dominio puede entenderse en un continuo generalidad-concreción; de forma que a mayor concreción, se entiende que existe una mayor calidad en la definición.

Como aspectos implicados en la definición están: *a)* exclusividad y *b)* exhaustividad. Así, entendemos que las unidades que definen el dominio no deben solaparse *(a)* y deben cubrir el dominio en su totalidad *(b)*.

Evidentemente, la definición del dominio depende, en gran medida, de las posibilidades del material instruccional. A mayor ambigüedad en el contenido de la instrucción, necesariamente habrá mayor generalidad en la definición del dominio. Así, cuando conocemos bien cuáles son los componentes y procesos subyacentes, el dominio puede ser bien especificado y, por tanto, bien medido; sin embargo, si ello no es posible, afecta directamente a todas las características e indicadores integrados en el proceso de medida.

Este aspecto se relaciona directamente con la validez de contenido de la prueba, dado que se trata de delimitar el objeto de la medida.

1.2. *Estructuración del dominio*

Además de la definición de las unidades que constituyen el dominio, se debe analizar y especificar su estructura. Así, entendemos que pueden identificarse dos sistemas de configuración estructural:

1. *De estructura implícita.*—Corresponderá a aquellos dominios cuyas unidades tienen una característica propia independiente de las demás unidades (como, por ejemplo, dificultad teórica o complejidad

[2] Desde nuestra concepción implicarán, asimismo, las estrategias cognitivas definidas experimentalmente subyacentes a la solución de las tareas que lo componen.

[3] Entendidas que las unidades se especifican en sí mismas o por el sistema concreto en que son generables o formulables.

cognitiva) o bien aquellos cuya estructura es consecuencia de las relaciones estructurales entre las unidades (como, por ejemplo, relevancia del contenido, nivel de generalidad, secuencia, etc.).

2. *De estructura resultante.*—Corresponderá a aquellos dominios en los que el análisis empírico de los comportamientos de los sujetos respecto a las unidades y sus relaciones impliquen consecuencias modificativas de la estructura del mismo (por ejemplo, dificultad empírica, etc.).

Por otra parte, puede considerarse que existen dos tipos de acercamiento para la determinación de la estructura del dominio:

a) *Juicio/objetivo.*—Basado en estrategias de jueces, se tratará de analizar las características del dominio, a partir del conocimiento que éstos posean del mismo. Hemos añadido el término «objetivo» dado que ello supone dinámicas bien definidas y análisis estadísticos oportunos para determinar la calidad y fiabilidad del trabajo de los jueces.
b) *Empírico/objetivo.*—Se tratará de comprobar características estructurales del dominio, a partir de datos resultantes de la aplicación de pruebas representativas del mismo.

En cualquier caso, entendemos que ambos planos de análisis no son excluyentes y que depende de las características del contenido del dominio, así como del objetivo del estudio a realizar, el uso de los mismos será independiente o interactivo.

Por último, respecto a la estructura, queremos señalar que en los TRC se ha venido considerando, a nivel teórico, que los dominios a medir eran homogéneos —desde nuestra posición, «sin estructura» o universos simples—; sin embargo, prácticamente todos los acercamientos han obviado esta consideración y han tratado dominios heterogéneos —estructurables, universos compuestos— sin hacer referencia a ello. Este hecho resulta curioso, dado que en la mayoría de capítulos metodológicos —determinación de estándares, fiabilidad, análisis de *items*, etc.— se trabaja con el supuesto implícito y/o explícito de la homogeneidad del dominio.

Desde un punto de vista operativo, se puede entender la existencia de dominios homogéneos como los subdominios de otro heterogéneo.

1.3. *El problema de la densidad diferencial dentro del dominio como universo de medida*

En el problema que vamos a tratar aquí, más que pretender ofrecer soluciones —aunque fueran tentativas— queremos aportar un punto de reflexión que hace referencia a la densidad diferencial respecto a las posibilidades métricas de subdominios. Entendemos por «densidad diferencial» el hecho de que ciertos objetivos presentan unas mayores posibilidades de medición que otros, generando un mayor número de elementos. Así, si se ha especificado un dominio educativo mediante cualquier sistema de generación, se podría conceptualizar éste como universo de medida, a partir del cual un TRC se entiende como una muestra representativa del mismo. En este sentido, el dominio como universo de medida definido provee el marco referencial necesario para efectuar este tratamiento. Sin embargo, es preciso señalar que dependiendo del sistema de especificación del dominio que se haya utilizado, el muestreo podría efectuarse de una u otra forma, considerando que el sistema de especificación generará universos de medida finitos (o definidos en todos y cada uno de sus elementos) o infinitos (definidos más o menos operativamente respecto a la forma de los elementos, pero no en su formulación individual).

Por otra parte, y con referencia a la constitución de universos finitos, es preciso señalar que si bien la propia definición del universo respecto a los elementos que especifican la medida de cada objetivo es diversa por su número de elementos, no debe plantearse establecer un isomorfismo directo entre la estructura del universo de medida y la estructura de la prueba. Esto es así dado que la posibilidad —mayor o menor— de generación de elementos no está relacionada directamente con la riqueza o importancia del contenido. La estructuración de la prueba debe responder a la estructuración jerárquica establecida por los evaluadores o bien por el juicio de los expertos consultados a tal efecto.

En cualquier caso, como veremos a lo largo de este trabajo, en muchas ocasiones la necesidad del propósito métrico impide que sean aplicables tecnologías que para otros menesteres han demostrado ser de gran utilidad [4], lo cual no es óbice para seguir persistiendo en la búsqueda de aproximaciones basadas en el rigor metodológico necesario que asegure la objetividad en los resultados.

2. Elementos en la definición del dominio

Como señalamos en el apartado anterior, entendemos por definición del dominio la especificación de sus unidades. En este sentido, es obvio que la ambigüedad y dispersión de la materia a definir son las variables que inicialmente van a determinar el alcance y precisión de dicha definición (Nitko, 1980, 1984). Con todo, aun a pesar de ser reduccionistas, estimamos que pueden identificarse tres niveles de definición: taxonómico, por objetivos y por *items* (o reglas de generación de *items*).

El nivel taxonómico puede desempeñar un doble papel: *a)* orientación del análisis de acercamiento al dominio, como una estructura apriorística de análisis y/o clasificación, y *b)* a partir de la comprobación empírica, como la síntesis de resultados genéricos descriptivos de la realidad. De este modo, se establece un *feed-back* teórico-empírico-teórico a través de los niveles inferiores de descripción (objetivos e *items*). El nivel taxonómico, si bien no es requisito ni fin de la medida, desempeña un papel de orientación en la identificación del universo de medida. Evidentemente, existen diversas orientaciones taxonómicas en educación [5] que pueden desarrollar este papel de orientación-guía en la identificación de los elementos del dominio (Fleishman y Quaintance, 1984).

El segundo nivel de trabajo lo constituye la definición por objetivos, los cuales implican en su propia definición un sentido finalista (Rodríguez Diéguez, 1979). No vamos a revisar aquí las diferentes definiciones de objetivos que se han propuesto (Popham y Baker, 1970; García Hoz, 1970;

[4] Por ejemplo, el muestreo de matrices es muy útil en aproximaciones de bancos de *items*, los cuales poseen una gran base normativa.

[5] Como la de Bloom y colaboradores (1956, 1969, 1971), Gagne (1971) o Scriven (1967), o como acercamiento de reformulación más funcional desde el *mastery learning*, como el «modelo de aprendizaje escolar» de Carroll (1963) y los trabajos de Bloom en esta línea (1968, 1971 *b*), entre otros.

Bloom *et al.*, 1969; Mager, 1973; Rodríguez Diéguez, 1979, entre otros); unicamente queremos señalar que suponen, a nivel de generalidad, la segunda unidad de definición [6] y se ajustan, en términos generales, a los requisitos señalados por Bradfield y Moredock (1957) respecto a las dimensiones de un fenómeno para que sea susceptible de medición.

En esta línea se ha desarrollado dentro de la ERC la experiencia más amplia, que ya hemos nombrado y comentaremos posteriormente, con el IOX de Popham como un centro de objetivos amplificados —de acuerdo con la tecnología y el término definidos por el propio autor— con utilidad para los profesionales de la educación para generar pruebas TRC.

Sin embargo, el nivel de especificación de los objetivos no es lo suficientemente concreto como para servir de unidad de definición. Unicamente se asumen cuando no es posible llegar a un mayor nivel de concreción. Se entiende, pues, que la definición más deseable para un dominio es la especificación de los *items* [7]. De este modo, se están desarrollando con gran ímpetu diversas opciones metodológicas y técnicas encaminadas a asegurar y aumentar el nivel de calidad de los elementos. Este punto estimamos que compete directamente a la medida, al ser los *items* las unidades mínimas de medición. Pasamos, pues, a revisar los procedimientos que se han propuesto y desarrollado en el marco de la ERC.

2.1. *Métodos de generación de items*

La formulación y escritura de *items* ha sido un tema que ha preocupado relativamente poco en los diversos modelos de medición, tanto en psicología como en educación. Sin embargo, con el desarrollo de las aproximaciones cognitivas al estudio de la inteligencia (Carroll, 1963, 1976, 1981; Sternberg, 1977, 1980, 1982, 1984) y su incorporación, al menos como línea de pensamiento, en los planteamientos de la llamada psicología de la instrucción (Glaser, 1976), con las repercusiones que ésta conlleva en los planteamientos del denominado *Mastery Learning* (Bloom, 1968, 1971, 1974; Bloom, Madaus y Hastings, 1981; Guskey, 1980 *a* y *b*, 1982, 1985), el estudio de los *items*, como componentes de medida que deben estar bien adaptados (y ser representativos de) al proceso de instrucción, ha tomado un gran interés. Así, se han realizado diversas revisiones respecto a la emergencia de la tecnología de escritura de *items* (Engel y Martuza, 1976; Berk, 1980; Roid y Haladyna, 1980 *a*, 1982, y Roid, 1984), así como se han editado algunos trabajos que pretenden servir de guía para mejorar la calidad de los *items* (Coffman, 1971; Conoley y O'Neil, 1979; Ebel, 1979; Gronlund, 1982; Roid y Haladyna, 1982). De todos ellos, quizá la revisión más completa del tema es la de Roid y Haladyna (1982) [8]. Sin embargo, en ella no aparece todavía una clasificación de métodos y técnicas de formulación de *items*.

Roid (1984, pág. 51) ofrece una clasificación que parte de tres categorías metodológicas: *a)* Procedimientos. *b)* Métodos basados en teorías. *c)* Métodos diagnósticos.

En la primera de ellas incluye los trabajos de Hively y colaboradores de 1968 respecto a los formatos de *items;* en la segunda, el diseño de facetas de Guttman (1959, 1970), y los *tests* basados en conceptos, reglas y principios de Markle y Tiemann (1970), y una tercera categoría, construcción de *items* basados en factores, en donde recoge los trabajos de Guilford (1967) y Meeker y Meeker (1975). Estos trabajos, desde nuestro punto de vista, no suponen necesariamente una técnica de escritura o formulación de *items* y ello nos hace pensar que quizá Roid (1984) ha intentado buscar un antecedente a planteamientos teóricos estructurales, como son el sistema LOGIQ o el IQI [9], y es posible que, por su parecido —sobre todo el LOGIQ— con los planteamientos formales de Guilford, asumiera los trabajos de éste como una tecnología de formulación de elementos. Finalmente, en la tercera categoría (diagnóstica) incluye los trabajos de Scandura (1970), Durnin y Scandura (1973), y diversas aproximaciones de base cognitiva, como Brown y Burton (1978), Glaser (1981) y Birenbaum y Tatsuoka (1982, 1983).

En este apartado realizamos una breve revisión de la tecnología de formulación de *items* siguiendo la clasificación de Roid (1984), la cual hemos modificado únicamente en el punto comentado con anterioridad —ver tabla 1—. Sin embargo, no vamos a extendernos en ellas por razones obvias de espacio, por lo que dejamos su análisis en profundidad para posteriores trabajos.

— PROCEDIMIENTOS.	— Formatos de *items*. — Transformaciones lingüísticas.
— METODOS BASADOS EN TEORIAS.	— Diseño de facetas. — Conceptos, reglas y principios: ● Aproximación Markle y Tiemann. ● Sistema LOGIQ. ● Inventario de Calidad Instruccional **(IQI).**
— PROCEDIMIENTOS ORIENTADOS AL DIAGNOSTICO.	— Algorítmico de Scandura. — Métodos basados en estrategias cognitivas.

TABLA 1: Clasificación de los procedimientos y métodos de formulación y escritura de *items* (basada en Roid, 1984).

2.1.1. Procedimientos

Los *formatos de items (items forms)* constituyen quizá la tecnología más conocida. Propuestos por Hively (1966) han tenido un amplio uso (Osburn, 1968; Freemer y Anastasio, 1969; Johnson, 1973; Vickers, 1973; Olympia, 1975; McClain, Wessels y Snado, 1975; Millman y Outlaw, 1978; Braby, Parrish, Guitard y Aagard, 1978), sobre todo en relación a elementos técnicos y cuantitativos. Respecto a su utilización para estructurar los contenidos no la detallamos, dado que son ampliamente conocidos y se incluyen en textos traducidos al castellano que han tenido mucha difusión, como es el caso de Popham (1980).

[6] Es evidente que dentro de los objetivos también puede considerarse un continuo de generalidad/operatividad, que se corresponde con ambigüedad/concreción de las unidades.

[7] O sistemas que permiten la generación de *items* con la misma estructura sintáctica, contenido equivalentes y, presumiblemente, el mismo nivel de dificultad.

[8] Este trabajo es un claro exponente de los puntos que acabamos de comentar.

[9] Los cuales no incluye en su trabajo de 1984, aunque reseña con extensión en Roid y Haladyna (1982).

Como ventajas de esta tecnología, de acuerdo con Roid y Haladyna (1982), tenemos las siguientes:

1. Pueden ahorrar tiempo de desarrollo del *test* si se realiza una inversión de tiempo antes, en la creación de tales formatos, en lugar de emplearlo en el desarrollo de cada *item* particular.
2. Definen bien los dominios que constituyen la base de los TRC, de modo que son interpretables sin ambigüedad al aportar una estimación insesgada de la proporción del dominio total que el estudiante domina.
3. Son especialmente aplicables en áreas cuantitativas complejas, como los problemas estadísticos de nivel superior o análisis de costos, en las que se pueden aprovechar las características de precisión y poder de cálculo de los ordenadores.
4. Se pueden realizar programas de ordenador para componer e imprimir aleatoriamente formas de pruebas. Así, no sólo se generan los ítems, sino también las formas de pruebas, eliminando una importante cantidad de tareas administrativas.
5. Son aplicables, en general, a todos aquellos campos en que se puede efectuar análisis de tareas o donde se pueda efectuar una definición jerarquizada del dominio instruccional.

Como inconvenientes, se pueden citar:

1. Su excesiva rigidez los hace difícilmente aplicables a las áreas o dominios cuya estructura y jerarquización no esté claramente explicitada.
2. Es un procedimientos sobre el que, no obstante su «edad» y el haber sido aplicado en múltiples ocasiones, prácticamente no existen estudios respecto a las condiciones para determinar las diversas partes que constituyen un formato de *items*, sus características métricas y las diversas opciones que se pueden dar junto con su repercusión.

Por otra parte, las *transformaciones lingüísticas* fueron propuestas por Bormuth (1970) como conexión lógica entre los *items* y el material del texto —libro/s de texto— (o, como él denomina, «de prosa»), dado que el aprendizaje se desarrolla normalmente tomando como referencia un libro de texto. Su propuesta se concreta en la transformación de frases o grupos de frases, de acuerdo con un procedimiento lógico que oriente tales tipos de modificaciones. En este sentido, Cronbach (1970) señala que este método debe restringirse a transformaciones simples, y Anderson (1972) presenta una técnica que emplea la paráfrasis para evaluar el nivel de comprensión (más que el de recuerdo).

Respecto a los procedimientos usados para este cometido, han sido diversos y se encuentran en las revisiones efectuadas por Roid *et al.* (1979), Roid y Haladyna (1982) y Roid (1984). Normalmente, incluyen una secuencia de cuatro fases, que se pueden sintetizar en las siguientes:

1. Búsqueda de frases clave relevantes en el proceso de instrucción.
2. Selección de las frases más relevantes.

3. Transformación de las frases —ver cuadro 1—.
4. Construcción de distractores para formatos de elección múltiple [10].

Frase ejemplo:

«Un disco magnético es un ejemplo de dispositivo de acceso directo no secuencial».

Palabra clave:

«Dispositivo de acceso directo».

Palabra transformada:

¿De qué es un ejemplo un disco magnético?
De un dispositivo de ———————— no secuencial.

a. Comunicación remota.
b. *Output* a los periféricos.
c. Acceso directo.
d. Edición de ficheros.

CUADRO 1: Ejemplo de procedimiento TP con palabras-clave excesivamente largas.

	Dinámico	*Estático*
Animado	Animal, hombre, insecto, Juan	
Simbólico	Película, juego, sonido	Libro, cuadro, carta
No simbólico	Viento, ruido, presión	Roca, casa, pala
Abstracto	Amor, esperanza, animal, hombre	Longitud, tamaño, kilos

CUADRO 2: Ejemplo de la utilización del sistema de categorización de Frederiksen (1975).

Los *items* resultantes de este tipo de transformaciones pueden ser de diversos tipos. Conoley y O'Neil (1979) proponen para su clasificación la taxonomía de Bloom, de forma que pueden identificarse *items* de acuerdo con cuatro niveles: conocimiento, comprensión, aplicación y análisis.

Las ventajas de usar TL se pueden sintetizar en los siguientes puntos:

1. Proporcionan un conjunto de procedimientos para desarrollar *tests* que estén lógicamente relacionados con los textos que se utilicen como material instruccional.
2. Los métodos definidos operacionalmente, se pueden comunicar con facilidad a otros investigadores y profesionales del ámbito de interés, de manera que se puedan diseñar pruebas que repercutan en el avance de las tecnologías de instrucción y evaluación.

[10] Para este cometido, Frederiksen (1975) propone un sistema de categorización para agrupar las palabras clave de acuerdo con algún principio genérico de similitud —ver cuadro 2—.

3. Las definiciones claras de las tareas cognitivas realizadas previamente a que el sujeto pase la prueba pueden ser de gran ayuda en el conocimiento de los procesos implicados en el aprendizaje a partir de un texto.
4. Se pueden computerizar, al menos en la forma de asistido por ordenador (Schulz, 1979), para ayudar en la búsqueda de palabras clave y su repetición en el texto. Se podrán realizar avances importantes con la utilización de tecnologías de tratamiento de textos y análisis de palabras sobre libros de texto completos.
5. Los avances en las relaciones interfrase e interpárrafo, basadas en el análisis lingüístico, pueden proporcionar nuevos métodos para explorar las habilidades de mayor nivel implicadas en el aprendizaje a partir de un texto instruccional.

Sus desventajas son:

1. No existe suficiente investigación realizada para garantizar una operatividad satisfactoria en todos los niveles de conocimiento. Por ello, los *tests* actuales presentan claramente el defecto de medir habilidades poco relevantes.
2. La mayor parte de los procedimientos propuestos son más «promesas» que «realidades» desde el punto de vista de su bagaje metodológico.
3. Requieren aún una gran cantidad de investigación lingüística y de procesamiento cognitivo de material escrito.
4. En el estado actual, buena parte de las técnicas son de difícil operacionalización, de forma que no garantizan su repetitividad objetiva a un nivel satisfactorio.
5. Asimismo, gran parte de las técnicas no son susceptibles de implementación por ordenador por la razón expresada en el punto anterior: su deficiente nivel de operacionalización que no permite determinar con claridad las reglas necesarias para su automatización. Ello conduce, indefectiblemente, a dificultar su utilización por la ingente labor que supone.

2.2. *Métodos basados en teorías*

La teoría estructural de facetas (o diseño de facetas) fue formulada por Guttman (1959) en el ámbito de la teoría de la personalidad y aplicada posteriormente por este mismo autor (1965, 1969) a los *tests* de rendimiento. El diseño de facetas aparece sobre la base de que «*... un constructo puede explicarse en términos de medida sobre una base ordenada y precisa...*», su esencia es que «*... cualquier constructo puede describirse en términos de sus componentes*» (Roid y Haladyna, 1982, pág. 127). Con el diseño de facetas: «*... Primero, se especifica un sistema definicional para el universo de contenido y observaciones en la forma de una sentencia directriz ('Mapping Sentence'). Segundo, se realizan las especificaciones acerca de las facetas de la sentencia directriz... Las definiciones y especificaciones conllevan una hipótesis estructural que se comprueba mediante datos empíricos*» (Guttman, 1969, pág. 56). Dentro de un diseño de facetas pueden identificarse los siguientes elementos: componentes del diseño y niveles funcionales de desarrollo.

a) Componentes de un diseño de facetas: sentencias directrices, facetas y elementos —ver cuadro 3—. Las sentencias directrices son expresiones resumidas que marcan la estructura (orden, componentes y subcomponentes) y límites de un constructo o dominio educativo (que en el caso de los *tests* de rendimiento puede ser un objetivo operativo). Dichas sentencias incluyen partes fijas —similares a los *item form*— y variables —facetas—, las cuales son series de reemplazamiento compuestas por diferentes variaciones de su contenido.

b) Niveles funcionales de desarrollo. Para este cometido existen algunas revisiones y consejos de gran utilidad: Engel y Martuza (1976), Runkel y McGrath (1972), Berk (1978) y Roid y Haladyna (1982). En todos ellos aparecen diversas recomendaciones respecto a la aplicación del diseño de facetas como un análisis de contenido que permite generar *items*, los cuales teóricamente presentan *a priori* los mismos niveles de dificultad y discriminación. Esto último no es totalmente cierto y depende de la calidad y extremado cuidado en la definición de la sentencia directriz (Jornet, 1987).

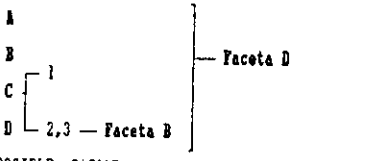
Este método permite, considerando elementos adicionales de definición (reglas de generación aleatoria y algoritmos para la generación de elementos) —ver cuadro 3—, construir para cada *item* un banco de generación automática por ordenador, de forma que convenientemente estructurados de acuerdo con el diseño de facetas pueden constituir los elementos básicos de construcción de un *test* administrado por ordenador (TAO-CAT). Esta metodología ha sido probada con éxito (Jornet, 1987) en relación a un curso de BASIC, como un instrumento de generación de *items*. Todo ello constituye, asimismo, un paso previo y un sistema que permite la determinación subsiguiente de las variables implicadas, sus relaciones estructurales y características métricas que posibilitan abordar satisfactoriamente una prueba del tipo CAT en sentido pleno.

Las ventajas que se señalan a este método, se pueden sintetizar en los siguientes puntos:

1. Facilita la creación sistemática de *stems*, *items* y distractores. Este aspecto es muy importante tenerlo en cuenta dada la dificultad de tal tipo de generaciones (Engel y Martuza, 1976; Roid y Haladyna, 1978, 1980*a*, 1982). Por otra parte, el carácter automático de la generación posibilita que ésta se realice por computador (Millman, 1980, 1984; Roid, 1984; Jornet, 1987).
2. Considerando la estructura de una sentencia directriz, se pueden crear múltiples formas paralelas a partir de la generación aleatoria de elementos, que lógicamente pueden considerarse de la misma dificultad y discriminación (Engel y Martuza, 1976; Roid y Haladyna, 1982).
3. Existe contigüidad entre los distractores, siendo unos más plausibles que otros. Así, el análisis de distractores conlleva valor diagnóstico, ya que añade matizaciones cualitativas en la valoración del producto. De este modo, Engel y Martuza (1976) señalan que «*... la selección consistente del tipo de distractores orienta hacia una estrategia específica para la recuperación*» (pág. 26).
4. Siendo que el diseño de facetas parte de objetivos, facilita estable-

DIAGRAMA DE GENERACION

- Asignaciones:



NUMERO DE ITEMS POSIBLE: $2^{*}C^{4*2}_{5}$ TIPO DE ITEMS: Elección Múltiple (4 alternativas)

MODULO ALGORITMICO DE GENERACION POR COMPUTADOR

- El presente módulo genera aleatoriamente los ítems pertenecientes a este objetivo, según las condiciones expuestas en el apartado anterior; así como, calcula la solución correcta y los distractores, ofreciendosela al evaluador.

```
 100   K4=0:FOR I=1 TO 5
 110   FOR J=1 TO 3
 120   K1=1:K2=FM(J)
 130   GOSUB 1000
 140   F(I,J)=K
 150   NEXT J
 160   IF F(I,3)=1 THEN GOTO 200
 170   K1=1:K2=FM(4)
 180   GOSUB 1000
 190   F(I,4)=K
 200   NEXT I
 210   FOR I=1 TO 5
 220   GOSUB 1100
 230   IF F(I,3)=1 THEN FV(F(I,1))=FM(F(I,1)):GOTO 260
 240   GOSUB 1300
 250   IF I=3 THEN K4=K4+1:GOSUB 2100
 260   NEXT I
 262   K1=0:K2=1
 264   GOSUB 1000
 266   IF K=0 THEN K1=1 ELSE FA(1)=1:K1=2:K3=1
 268   FOR I=K1 TO 3
 270   K1=2:K2=7
 272   GOSUB 1000:FA(I)=K
 274   FOR J=1 TO I-1
 276   IF FA(J)=K THEN GOTO 270
 277   NEXT J
 278   NEXT I
 280   K1=1:K2=3
 282   GOSUB 1000
 284   IF K3=1 THEN K2=FA(K):FA(K)=FA(1):FA(1)=K2:K3=K
 300   PRINT "Dadas las siguientes asignaciones a las  variables, cuyos valores iniciales
       son ";
 310   IF F(0,1)<>0 THEN PRINT "A=4, B=6 y C=5, " ELSE  PRINT "todos cero, "
 320   FOR I=1 TO 5
 330   PRINT I*10;" LET ";A$(F(I,1));B$(F(I,2));
 340   IF F(I,3)=1 THEN PRINT:GOTO 360
 350   PRINT C$(F(I,3));B$(F(I,4))
 360   NEXT I
 370   PRINT  "señalar de entre las siguientes afirmaciones  cuál refleja el estado final
       de las variables después  de su ejecución por el ordenador: "
 380   PRINT TAB(10) "OPCIONES: "
 390   PRINT TAB(21) "A     B     C "
 400   FOR I=1 TO 3:GOSUB 2000
 410   PRINT TAB(16) "-" TAB(20) S1 TAB(26) S2 TAB(32) S3 TAB(40) E$(I)
```

CUADRO 3(Cont.): Ejemplo de Sentencia Directriz y su desarrollo para Generación Automática de Items (JORNET, 1987).

OBJETIVO: Dadas las asignaciones a tres variables numéricas en forma de un programa de 5 pasos el alumno reconocerá cuál será el estado final de las mismas trás su ejecución por el ordenador.

SENTENCIA DIRECTRIZ

"Dadas las siguientes asignaciones a las variables, cuyos valores iniciales son **Faceta D**,

5 asignaciones:
1) Faceta A Faceta B Faceta C (Faceta B)
2) Faceta A Faceta B Faceta C (Faceta B)
3) Faceta A Faceta B Faceta C (Faceta B)
4) Faceta A Faceta B Faceta C (Faceta B)
5) Faceta A Faceta B Faceta C (Faceta B)

señalar de entre las siguientes afirmaciones cuál refleja el estado final de las variables después de su ejecución por el ordenador:

OPCIONES:

A B C

a)
b)
c)
(ver algoritmo de generación)

d) - Ninguna de las anteriores

FACETA A. Variables.	FACETA B. Valores.		FACETA C.(*) Operadores Aritméticos.	Faceta D. Valores iniciales.
1. A=	1. 0	5. A	1.	1. A=4, B=6 y C=5
2. B=	2. 1	6. B	2. +	2. todos cero
3. C=	3. 2	7. C	3. -	
	4. 3		4. *	

(*) No se incluye el operador (/), dado que supondría un incremento notable en la dificultad de los ítems.

ALGORITMOS DE GENERACION

1. Asignaciones. 1.1. Se generan aleatoriamente con reemplazamiento los elementos de las Facetas A, B y C. 2.2. Se genera el elemento de la Faceta D condicionado a: 2.2.1. En el caso en que dos de las tres primeras asignaciones contengan asignaciones 0 a las variables, se otorga elemento 1. 2.2.2. En el caso en que 2.1. no se de, se otorga elemento 2.

2. Opciones.
a, b y c. Se extraen aleatoriamente y sin reemplazamiento de los siguientes parámetros:

	A	B	C	V=Valor correcto
1	V	V	V	
2	V+1	V	V	
3	V-1	V	V	
4	V	V+1	V	
5	V	V-1	V	
6	V	V	V+1	
7	V	V	V-1	

d. Es constante

Si al extraer aleatoriamente el elemento de la Faceta C sale el 1, no se extrae por segunda vez un elemento de la Faceta B, ya que al no haber operador se trata de una asignación directa.

CUADRO 3: Ejemplo de Sentencia Directriz y su desarrollo para Generación Automática de Items (JORNET, 1987).

```
420   NEXT I
430   PRINT TAB(10) "- Ninguna de las anteriores "  TAB(40)  "d"
440   PRINT "La solución correcta es: ";
450   IF E5<>0 THEN PRINT E$(E) ELSE PRINT "(d)"
460   END
1000  K=X1+(X2-X1)*RND(1)
1010  IF K-INT(K)>0.5 THEN K=INT(K)+1 ELSE K=INT(K)
1020  RETURN
1100  ON  F(I,2) GOTO 1110,1120,1130,1140,1150,1160,1170
1110  FM(F(I,1))=0:GOTO 1180
1120  FM(F(I,1))=1:GOTO 1180
1130  FM(F(I,1))=2:GOTO 1180
1140  FM(F(I,1))=3:GOTO 1180
1150  FM(F(I,1))=FV(1):GOTO 1180
1160  FM(F(I,1))=FV(2):GOTO 1180
1170  FM(F(I,1))=FV(3)
1180  RETURN
1300  ON F(I,3)-1 GOSUB 1400,1600,1800
1310  RETURN
1400  ON  F(I,4)  GOTO 1405,1410,1420,1430,1440,1450,1460
1405  FV(F(I,1))=FM(F(I,1))+0:GOTO 1470
1410  FV(F(I,1))=FM(F(I,1))+1:GOTO 1470
1420  FV(F(I,1))=FM(F(I,1))+2:GOTO 1470
1430  FV(F(I,1))=FM(F(I,1))+3:GOTO 1470
1440  FV(F(I,1))=FM(F(I,1))+FV(1):GOTO 1470
1450  FV(F(I,1))=FM(F(I,1))+FV(2):GOTO 1470
1460  FV(F(I,1))=FM(F(I,1))+FV(3)
1470  RETURN
1600  ON  F(I,4)  GOTO 1605,1610,1620,1630,1640,1650,1660
1605  FV(F(I,1))=FM(F(I,1))-0:GOTO 1670
1610  FV(F(I,1))=FM(F(I,1))-1:GOTO 1670
1620  FV(F(I,1))=FM(F(I,1))-2:GOTO 1670
1630  FV(F(I,1))=FM(F(I,1))-3:GOTO 1670
1640  FV(F(I,1))=FM(F(I,1))-FV(1):GOTO 1670
1650  FV(F(I,1))=FM(F(I,1))-FV(2):GOTO 1670
1660  FV(F(I,1))=FM(F(I,1))-FV(3)
1670  RETURN
1800  OF  F(I,4)  GOTO 1810,1815,1820,1830,1840,1850,1860
1810  FV(F(I,1))=0:GOTO 1870
1815  FV(F(I,1))=FM(F(I,1))*1:GOTO 1870
1820  FV(F(I,1))=FM(F(I,1))*2:GOTO 1870
1830  FV(F(I,1))=FM(F(I,1))*3:GOTO 1870
1840  FV(F(I,1))=FM(F(I,1))*FV(1):GOTO 1870
1850  FV(F(I,1))=FM(F(I,1))*FV(2):GOTO 1870
1860  FV(F(I,1))=FM(F(I,1))*FV(3)
1870  RETURN
2000  ON  FA(I)  GOTO  2010,2020,2030,2040,2050,2060,2070
2010  S1=FV(1):S2=FV(2):S3=FV(3):GOTO 2080
2020  S1=FV(1)+1:S2=FV(2):S3=FV(3):GOTO 2080
2030  S1=FV(1)-1:S2=FV(2):S3=FV(3):GOTO 2080
2040  S1=FV(1):S2=FV(2)+1:S3=FV(3):GOTO 2080
2050  S1=FV(1):S2=FV(2)-1:S3=FV(3):GOTO 2080
2060  S1=FV(1):S2=FV(2):S3=FV(3)+1:GOTO 2080
2070  S1=FV(1):S2=FV(2):S3=FV(3)-1
2080  RETURN
2100  IF K4=4 THEN 100
2110  IF  ((FV(1)=0)  AND  (FV(2)=0)) OR ((FV(1)=0) AND  (FV(3)=0))  OR  ((FV(2)=0)  AND
      (FV(3)=0)) THEN FV(1)=4:FV(2)=6:FV(3)=5:I=1:F(0,1)=1:RETURN 210
2120  RETURN
```

CUADRO 3(Cont.): Ejemplo de Sentencia Directriz y su desarrollo para Generación Automática de Items (JORNET, 1987).

EJEMPLO DE ITEM

Dadas las siguientes asignaciones a las variables, cuyos valores iniciales son todos cero,

```
10 LET  B=2
20 LET  A=3
30 LET  C=0
40 LET  C=A*B
50 LET  A=C-B
```

señalar de entre las siguientes afirmaciones cuál refleja el estado final de las variables después de su ejecución por el ordenador:

OPCIONES:

```
        A   B   C

  -     4   2   6 ..................  a
  -     4   2   7 ..................  b
  -     3   2   6 ..................  c
  -  Ninguna de las anteriores ........  d

  La solución correcta es: (a)
```

CUADRO 3(Cont.): Ejemplo de Sentencia Directriz y su desarrollo para Generación Automática de Items (JORNET, 1987).

cer una relación lógica entre el contenido de la instrucción y los *items* de los *tests*.

Por otra parte, las limitaciones imputadas son:

1. Es difícil su aplicación sobre material poco estructurado. De este modo, es útil en materias científicas (como matemáticas, física, etcétera); sin embargo, es poco factible su aplicación sobre lenguaje, ciencias sociales, etc. (Martuza, 1979; Roid, 1984).
2. Pueden encontrarse diferencias entre los jueces al formular sentencias directrices, facetas y elementos (Engel y Martuza, 1976; Roid *et al.*, 1979, 1980; Martuza, 1979). Ellas, en nuestra opinión, se darán con mayor facilidad en tanto en cuanto el material a estructurar sea más ambiguo.
3. El desarrollo de un diseño de facetas es muy costoso.
4. Existe todavía poca investigación empírica respecto a la viabilidad de los diseños de facetas.

En síntesis, si bien es una tecnología que se conoce aún poco, parece obvio que facilita en gran medida la estructuración y especificación de un dominio, tanto por la lógica interna que aporta como por evitar al constructor de la prueba el tener que formular uno a uno todos los elementos posibles; es más, considerando que de cada sentencia directriz se genera más de un elemento, presumiblemente de la misma dificultad y discriminación, facilita también la extracción de formas paralelas (clásicas o aleatorias).

Concepto de Gramática: Antónimo

Una palabra que:

Atributos Críticos:
1. Tiene un significado opuesto al significado de otra palabra.
2. Pertenece a la misma categoría gramatical que la otra palabra.
3. Es una palabra totalmente diferente, no un derivado de otra palabra.

Atributos Variables:
4. Puede tomarse de varias categorías gramaticales.
   a) Nombres c) Pronombres e) Adjetivos
   b) Verbos d) Adverbios f) Preposiciones
5. La longitud silábica relativa de las dos palabras puede ser:
   a) Igual
   b) Desigual
6. Puede existir oposición del significado:
   a) a lo largo de algún continuo
   b) en un sentido dicotómico

Ejemplos para la Instrucción
1. Malo; Bueno 4e, 5a, 6a
2. Peligro; Seguridad 4a, 5b, 6a
3. Vivir; Morir 4b, 5a, 6b
4. El; Ella 4c, 5b, 6b
5. Rápidamente; Lentamente 4d, 5b, 6a
6. Con; Contra 4f, 5b, 6b

Contraejemplos para la Instrucción
1. Vanidoso; Codicioso ..... no cumple AC1
2. Razón; Motivo ..... no cumple AC1
3. Nosotros; Vosotros ..... no cumple AC1
4. Arriba; Sobre ..... no cumple AC1
5. Alegremente; Triste ..... no cumple AC2
6. Feliz; Infeliz ..... no cumple AC3
7. Capaz; Incapaz ..... no cumple AC3
8. Discutible; Acuerdo ..... no cumple AC2

Ejemplos para la Prueba
1. Caliente; Frío 4e, 5b, 6a
2. Pérdida; Ganancia 4a, 5a, 6a
3. Subir; Bajar 4b, 5b, 6b
4. Tú; Yo 4c, 5a, 6b
5. Alegremente; Tristemente 4d, 5b, 6a
6. Sobre; Bajo 4f, 5a, 6b

Contraejemplos para la Prueba
1. Imaginario; Fantástico ..... no cumple AC1
2. Silla; Sofá ..... no cumple AC1
3. Antes; A continuación ..... no cumple AC1
4. Lóbrego; Brillar ..... no cumple AC2
5. Afinado; Desafinado ..... no cumple AC3
6. Válido; Inválido ..... no cumple AC3
7. Débil; Fuertemente ..... no cumple AC2
8. Cantar; Silencio ..... no cumple AC2

Muestra de Item del test
¿Cuál de los siguientes pares de palabras son antónimos?
(a) Imaginario - Fantástico
(b) Elevar - Bajar
(c) Válido - Inválido
(d) Débil - Fuertemente

Repuesta Correcta (b)

CUADRO 4: Ejemplo de un Análisis de Concepto para desarrollar TRCs de Aprendizaje de Conceptos. Adaptado de TIEMANN y MARKLE (1978a)

Por otra parte, la aproximación de Tiemann y Markle supone un método de generación de elementos basados en la utilización combinada de ejemplos y contra-ejemplos de un determinado concepto. Para ello parten de una definición de «concepto» muy útil, en los que identifican atributos críticos y variables —según su terminología—, de forma que la adquisición de conceptos se evalúa a partir de dos estrategias: *a)* comprobando la generalización en base a la correcta aplicación del concepto a un miembro de la clase conceptual no utilizado previamente, y *b)* comprobando la discriminación de algo como no perteneciente a la clase conceptual estudiada. Por otra parte, Tiemann y Markle (1978 *a* y *b*, 1983) indican que también es posible elaborar *items* que midan al mismo tiempo generalización y discriminación, como en el caso del ejemplo que adjuntamos —ver cuadro 4—.

En segundo lugar, el sistema LOGIQ es un método de clasificación que puede tomarse como medio útil para categorizar objetivos e *items*, así como para planificar y orientar la producción de estos últimos. Basado en el trabajo de Miller y colaboradores (1973, 1978, 1979), está estrechamente relacionado con la metodología IQI (Merrill *et al.*, 1979) por las dimensiones de contenido de los hechos, conceptos y principios que puede encontrarse en ambos sistemas. Asume tres dimensiones —ver figura 1—:

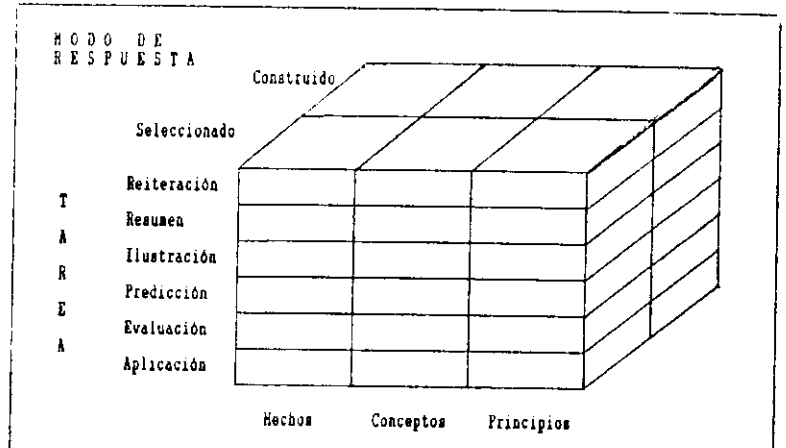


FIGURA 1: Diagrama de la tipología de clasificación de objetivos e items de tests del Sistema LOGIQ

a) *Contenidos.*—Define objetos, clasificaciones de los mismos y sus relaciones para concretar el dominio. Incluye las siguientes categorías: hechos, conceptos y principios.
b) *Tarea.*—Operación intelectual referida a la forma en que se utiliza el contenido. Incluye las siguientes subdimensiones: reiteración, resumen, ilustración, predicción, evaluación y aplicación.

c) *Modo de respuesta.*—Formas en que responde el sujeto ante la situación definida por las otras dos dimensiones. Incluye dos categorías: respuesta seleccionada y respuesta construida.

La generación de elementos se ordena de acuerdo con la confluencia de las categorías pertenecientes a las tres dimensiones definidas por este sistema.

Por último, el IQI es un sistema que orienta la evaluación de la consistencia y adecuación de los objetivos, materiales instruccionales e *items*. Se ha desarrollado para el entrenamiento militar al objeto de aportar una orientación, tanto para los que desarrollan los sistemas instruccionales como para los propios instructores, que tienen generalmente una formación deficiente en aspectos docentes y de manejo de pruebas. No obstante, este sistema puede ser también muy útil para los diseñadores de sistemas instruccionales y de pruebas en todos los niveles. El IQI es un subsistema del modelo de Desarrollo de Sistemas Instruccionales (DSI), que ha sido adoptado, en todas las ramas del Ejército de los Estados Unidos, como el método más adecuado para el desarrollo de los cursos de entrenamiento.

En el desarrollo del sistema IQI pueden identificarse dos etapas: la primera de ellas ligada al trabajo de Merrill y colaboradores, que cubre desde 1973 a 1978, y la segunda encabezada por Ellis y Wullfeck, antiguos colaboradores de Merrill, desde 1979. Sin embargo, pese a la identificación de estas etapas, vamos a exponer el estado actual del sistema IQI, obviando las diferencias entre ambas etapas por considerarlas normales en la evolución de un procedimiento que se está aplicando y estudiando constantemente.

El IQI se caracteriza por analizar la situación de instrucción en base a un sistema bidimensional en el que se definen tareas y contenidos —ver cuadro 5—.

Especificar los pasos concretos de desarrollo de la evaluación en este sistema sería sumamente extenso y sobrepasaría con mucho los límites de este trabajo. Así, remitimos para ello a la exposición-revisión de Merrill, Reigeluth y Faust (1979) o la de Roid y Haladyna (1982), o bien a las sistematizaciones del método de Wulfeck y Ellis (1981) o de Ellis y Wulfeck (1982).

Respecto a las ventajas e inconvenientes que presentan estas aproximaciones, se pueden extraer las siguientes notas:

En relación a la aproximación de Markle y Tiemann, Roid (1984) señala que su utilización es muy compleja y requiere un amplio período de aprendizaje, así como, si es posible, ayuda específica al evaluador.

Por otra parte, desde nuestro punto de vista, además de ser compleja, está orientada a lo que podríamos denominar procesos «superiores» de pensamiento. Así, no ha tenido otra aplicación que la que le han dado los autores del método. No obstante, nos parece una aproximación sumamente interesante y que requiere una amplia base experimental, la cual debería enraizarse en temas básicos de la psicología cognitiva, como es la investigación en resolución de problemas y razonamiento (Newell, 1980; Simon y Reed, 1976), razonamiento deductivo (Rips y Markus, 1977; Johnson-Laird y Steedman, 1978; Evans, 1982; Johnson-Laird, 1983) y razonamiento inductivo (Tversky y Kahneman, 1974 y 1978; Gigk y Holyoaka, 1980).

TAREA	CONTENIDOS				
	HECHOS Recordar o reconocer Nombres, Fechas, Lugares, etc.	CATEGORIA Recordar características o clasificar objetos, sucesos o ideas de acuerdo con características.	PROCEDIMIENTO Secuencia de pasos recordada o utilizada en una única situación o en una única parte del equipo.	REGLA Recuerdo o uso de una secuencia de pasos, que se aplican a través de situaciones o a través de equipos.	PRINCIPIO Recordar, Interpretar o predecir por qué o cómo sucederá algo o las relaciones causa-efecto.
RECORDAR Recuperar o reconocer hechos, definiciones de conceptos, procedimientos, reglas, afirmaciones de principios.					
APLICACIÓN SIN AYUDA Tareas que requieren clasificar, ejecutar un procedimiento, utilizar una regla, explicar o predecir sin ningún tipo de ayuda excepto la memoria.					
APLICACIÓN CON AYUDA Lo mismo que el anterior, pero disponiendo de ayudas (instrucciones, etc.).					

Matriz Contenido * Tarea del Sistema IQI.

PROPOSITO DE LA INSTRUCCION	OBJETIVOS DE LA INSTRUCCION	TEST O PRUEBA DE LA INSTRUCCION		PRESENTACIONES DE LA INSTRUCCION
1. ¿Es Consistente?	2. ¿Es adecuado?	3. ¿Es consistente?	4. ¿Es adecuado?	5. ¿Es consistente?	6. ¿Es adecuado?
a)Clasificar el Propósito o nivel de Tareas, sobre la base de la orientación y requerimientos de transferencia de la lección. b)Clasificar los objetivos a nivel de Tareas mediante la inspección. c)Comparar los niveles de Tareas de ambos para asegurarse que son los mismos. d)Asegurarse de que los objetivos están completos.	a)Asegurarse de que están claramente especificados: .Conductas .Condiciones y, .Estándares debiendo incluir: .Objetos .Herramientas y restricciones .Capacidad	a)Clasificar los objetivos de acuerdo con la matriz de Tareas' Contenidos. b)Clasificar cada ítem de la prueba de acuerdo con la matriz de Tareas' Contenidos. c)Comparar las clasificaciones de cada ítem con su objetivo y asegurarse de que son las mismas. d)Asegurarse de que se miden todos los objetivos.	a)Asegurar que la fiabilidad de los ítems sea buena. b)Asegurar que el formato de cada ítem sea técnicamente correcto. c)Asegurar que otros aspectos de la prueba sean adecuados; Nivel de las respuestas, Forma input-output, Criterio para determinar el tiempo para las respuestas, Feed-back y mensajes, muestreo, secuenciación y nivel en el criterio.	a)Determinar el nivel de la tarea de los ítems pertenecientes a un objetivo. b)Determinar los formatos de presentación primaria que se precisan para la instrucción en dicho nivel de la tarea. c)Asegurarse de que dichos formatos de presentación primaria se utilicen en la instrucción y no otros.	a)Asegurarse de que los componentes de estrategia necesarios están incluidos en c con cada formato de presentación primaria. b)Asegurarse de que cada componente de estrategia posee las características necesarias.

CUADRO 5: Un resumen de los aspectos de calidad instruccional analizados por el sistema IQI (tomado de ROID y HALADYNA, 1982).

Respecto al sistema LOGIQ, Williams y Haladyna (1982) informan que si bien existe todavía poca investigación al respecto, ésta es muy prometedora. Sin embargo, para que este sistema sea realmente operativo se necesita todavía mayor investigación.

Desde nuestro punto de vista, nos parece un sistema bastante comprehensivo, si bien puede presentar problemas similares a los que encontró Guilford (1956, 1967, 1981) en el ámbito de la inteligencia, de forma que no todas las categorías puedan ser identificables al mismo nivel. En cualquier caso, el intento es atractivo y puede favorecer una estructuración más que de la formulación de *items*, de su identificación como unidades del dominio.

El sistema IQI quizá ha sido el más utilizado de los que hemos revisado en el apartado de métodos basados en teorías. Esto ha sido así dado que su marco de aplicación (instrucción en la Armada de los Estados Unidos) ha favorecido que se desarrollara una gran base de investigación, tanto básica como aplicada; sin embargo, este punto también conlleva desventajas. Así, está muy restringido en su aplicación a cuestiones militares (Roid y Haladyna, 1982; Roid, 1984) y los textos disponibles, en donde se recogen los métodos, al estar editados por estamentos militares, son poco accesibles.

Con todo, el IQI es un sistema extremadamente comprehensivo y puede integrar gran cantidad de otros métodos para propósitos específicos (Ellis y Wulfeck, 1980; Wulfeck y Ellis, 1981; Montague, Ellis y Wulfeck, 1983). En nuestra opinión, sería sumamente interesante investigar la aplicabilidad y desarrollo del sistema IQI en otros ámbitos educativos.

2.3. *Procedimientos orientados al diagnóstico*

La aproximación estructural o algorítmica de Scandura (1970, 1977; Durnin y Scandura, 1973) propone analizar las conductas implicadas en la solución de un problema en forma de un diagrama de flujo similar a los algoritmos de computador.

Esta propuesta incluye diversas implicaciones, tanto para la definición de ítems como para su construcción. Es más, de acuerdo con Béjar (1983), «*Scandura ha propuesto no tanto un algoritmo para generar un conjunto de* ítems, *como una teoría completa de* tests *de ejecución que conecta consideraciones instruccionales y de medición en un esquema unificado para* tests *de rendimiento*» (pág. 14). Ciertamente, su aproximación comprehensiva implica una visión dinámica de los *tests* de rendimiento, los cuales son consecuencia del análisis del proceso instruccional, que intentan reflejar fielmente.

Así, como señalan Scandura (1970, 1977) y Scandura y Durnin (1978), no sólo deben describirse las conductas propias de un estudiante respecto a una realización concreta, sino también deben especificarse, pues, las «reglas de competencia», es decir, el proceso que hace resolver correctamente un problema. Una ventaja de este método es que una vez se han especificado los algoritmos (o reglas de competencia), éstos se pueden usar para desarrollar el conjunto de *items*, el cual estará totalmente ajustado a la instrucción y permitirá adicionalmente analizar las elecciones incorrectas y errores de los estudiantes para orientar específicamente el proceso de recuperación.

Como es obvio, esta aproximación es útil con materiales bien estruc-

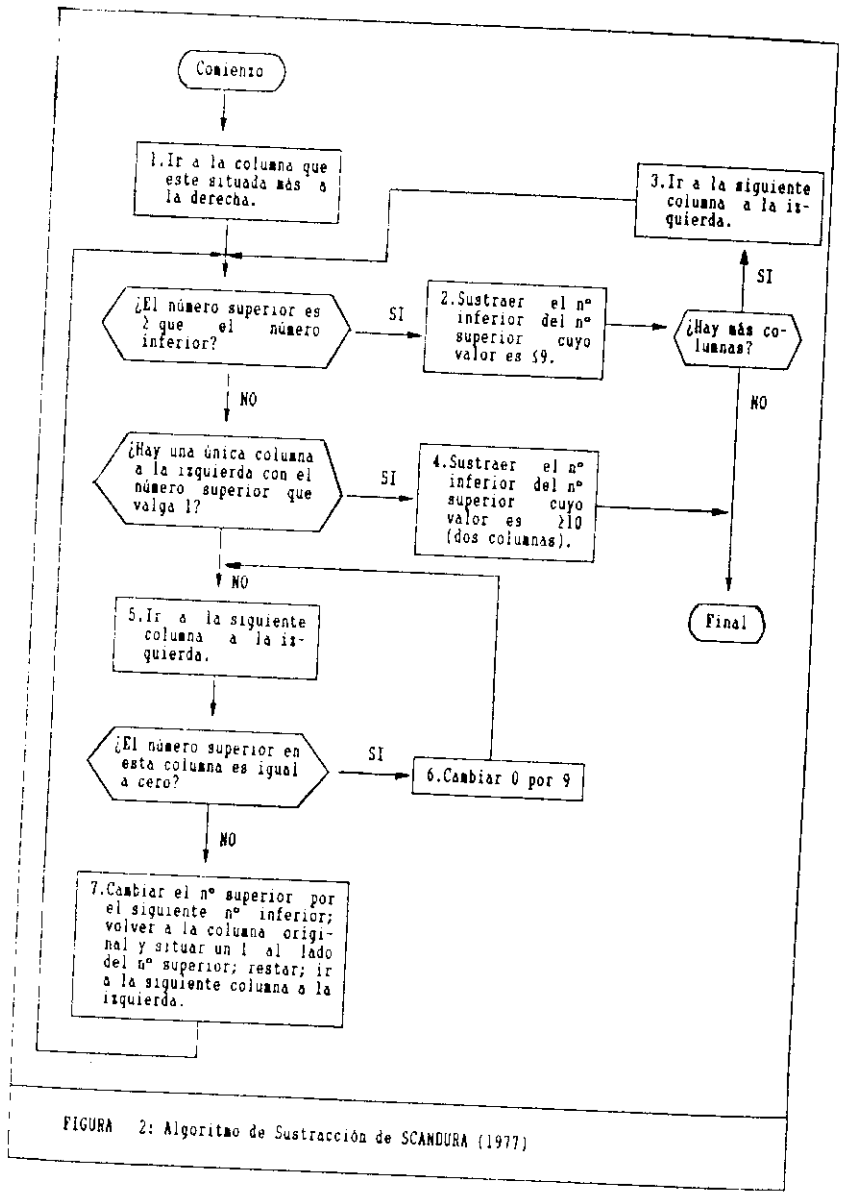


FIGURA 2: Algoritmo de Sustracción de SCANDURA (1977)

turados (o estructurables); así, como señala Roid (1984), es particularmente útil para problemas de matemáticas y áreas cuantitativas de las ciencias naturales o sociales.

Incluimos como ejemplo un algoritmo de Scandura (1977) realizado para explicar el proceso de sustracción —ver figura 2—.

Por otra parte, Roid (1984) revisa algunas aproximaciones que pretenden construir los *tests* desde el análisis de las estrategias cognitivas implicadas en la solución de problemas. Estos diseños de *tests* tienen en común el hecho de que al considerar los componentes y etapas en la solución de un problema, pueden incluir *subtests* o puntuaciones independientes para cada uno de ellos. Este aspecto, sin duda, conlleva grandes ventajas a nivel diagnóstico, dado que favorece el estudio de los errores. Sin embargo, como señala este autor, dado el escaso tiempo de desarrollo del tema, no existe todavía una categorización formal del mismo.

Así, reseña tres aproximaciones: Brown y Burton (1978), Birenbaum y Tatsuoka (1983) y Curtis y Glaser (1983). Las dos primeras se refieren a pruebas de aritmética, y la tercera, a pruebas de lectura.

En primer lugar, la aproximación de Brown y Burton (1978) se basa en un proceso usual para los programadores de ordenadores, el *debugging*, es decir, la ejecución paso a paso. Así, basándose en modelos de procesamiento de la información, pretenden determinar por el método descrito cuáles son las operaciones y problemas que usa comúnmente un estudiante en la solución de problemas de adición y sustracción. Ello está especialmente orientado a facilitar la identificación y el estudio de los errores.

Por otra parte, el trabajo de Birenbaum y Tatsuoka (1983), si bien no está orientado directamente hacia la tecnología de escritura de *items* [11], conlleva necesariamente el análisis de las estrategias cognitivas implicadas en la solución de un problema aritmético. Así, intenta mediante una aproximación algorítmica determinar los componentes cognitivos subyacentes en un *item*.

Por último, el trabajo de Curtis y Glaser (1983) se presenta en relación a la habilidad lectora, en la cual identifican cuatro áreas de funcionamiento cognitivo que conllevan implicaciones en el diseño de los *items: a)* decodificación de palabras, *b)* acceso semántico, *c)* procesamiento de frases y *d)* análisis del discurso. Existen, sin embargo, otras aproximaciones en este ámbito que todavía no se han desarrollado suficientemente, como las de Hoepfner (1978 o la de Drum, Calfee y Cook (1980). Esta última plantea como características a analizar en el proceso de lectura variables más superficiales que las de Curtis y Glaser (1983), como, por ejemplo, la longitud de la frase, la complejidad sintáctica o la plausibilidad de los distractores.

En cualquier caso, entendemos que todas las aproximaciones son excesivamente puntuales y que Roid (1984) quizá las clasifica precipitadamente. Así, ligadas al desarrollo de la psicología cognitiva y funcionalmente implicadas en todo el movimiento del diseño instruccional (Glaser, 1976), existen diferentes aproximaciones (Carroll, 1976; Sternberg, 1977, 1979) que conllevan indudables consecuencias respecto al diseño de *tests*

[11] Esta aproximación se presenta como un análisis de la congruencia de la valoración de un *item*. Así, descomponen el contenido del mismo en acierto (con estrategia correcta), acierto (con estrategia errónea) y error. Unicamente, la primera categoría se valora positivamente. Las otras dos se asumen como errores. La determinación de estrategias se realiza mediante un proceso empírico en el que se asocien estrategias/*items*. Birenbaum y Tatsuoka (1983) informan que este sistema de puntuación incrementa la consistencia interna y simplifica la estructura factorial de los *items*.

en su totalidad (Embretson, 1985) y, naturalmente, en la tecnología de escritura de *ítems*.

Las ventajas y limitaciones de estas aproximaciones se pueden sintetizar en los siguientes puntos. En primer lugar, en relación a la aproximación estructural de Scandura (1970, 1977), su aportación mayor, real y potencial, es la visión procesual que implica. De acuerdo con Béjar (1983), esta visión contrasta en gran medida con la visión estática de los *mastery tests* (por ejemplo, Block, 1971). Así, en general, con cualquier otro acercamiento el objetivo es determinar si el estudiante puede resolver un número (acordado) o proporción de *ítems* de un universo. Normalmente, esta proporción, a falta de una teoría, se establece por juicio (Béjar, 1983). Las ventajas de la aproximación de Scandura es que al determinar las reglas de competencia, se establecen, simultáneamente, los objetivos instruccionales y las reglas de generación de elementos, con lo que se facilita la igualación entre instrucción y universo de *ítems*. Por otra parte, como señalan Durnin y Scandura (1973), esta aproximación algorítmica presenta un gran potencial diagnóstico, dada la estructura que posibilita el análisis de los errores de los estudiantes, y, en concreto, estas características convierten este método en preferiblee frente al de formato de *ítems*.

Desde nuestro punto de vista, este acercamiento presenta indudables posibilidades, las cuales están todavía sin desarrollar, como son el análisis de estrategias de solución de problemas y su relación con niveles de éxito, enraizando dicho análisis en un acercamiento cognitivo riguroso. Podrían definirse los algoritmos como una consecuencia del análisis diferencial entre alumnos exitosos y no exitosos [12], o bien podría analizarse las diferencias de estrategia entre ellos a partir de estrategias prefijadas.

Por otra parte, respecto a los métodos basados en estrategias cognitivas, desde nuestro punto de vista sería necesario un planteamiento más comprehensivo, dado que existe un gran bagaje de investigación en la psicología cognitiva y que se traduce en nuevos conceptos respecto al diseño de *tests* (Embretson, 1985), que podrían reconvertirse con gran utilidad para la comprensión de los procesos de adquisición. En este sentido, estas aproximaciones constituyen, sin duda alguna, la visión más adecuada para llenar de contenido procesual las interpretaciones de las ejecuciones de los sujetos como productos en el *test*.

3. Elementos de estructuración del dominio

En este apartado vamos a revisar los acercamientos que se han propuesto para analizar la estructura del dominio.

Como señalamos anteriormente, entendemos que hay dos formas de abordar la estructuración de un dominio: *a)* estructura implícita y *b)* estructura resultante. Según esta diferenciación, pocas han sido las propuestas para el caso *(a)*. Para éste se entiende que existen variables internas al dominio (relevancia y afinidad temáticas, complejidad cognitiva o secuencialización, etc.) que suponen una estructura diferencial del mismo. Su determinación, evidentemente, debe realizarse sobre la base de análisis lógicos [13] y, necesariamente, el estudio de la estructura implícita debería realizarse a partir de aproximaciones basadas en el *scaling*. Así, la única propuesta que hemos encontrado a tal efecto es la de Hambleton (1984), respecto a la utilización del escalograma de Guttman para el análisis de la validez de constructo.

[12] En un estudio realizado con universitarios (y en otro contexto) encontramos que las diferencias en rendimiento se relacionaban con diferencias en estrategias cognitivas, las cuales se estructuraban en un continuo rigidez-flexibilidad (Jornet, 1980; Jornet y Suárez, 1984).

En nuestro caso, y en relación a un dominio genérico compuesto por los objetivos de un curso de BASIC —en conjunto, 58 objetivos— (Jornet, 1987), hemos comprobado una estrategia de análisis de aplicación con jueces basada en el escalamiento multidimensional, utilizable para estudiar la estructura implícita ligada a relevancia y complejidad teóricas.

Así, comprobamos dos estrategias de trabajo con jueces: molar —valorando comparativamente todos los objetivos del dominio— y molecular —basada en dos etapas—. La primera de ellas no ofreció resultados interpretables, y la segunda presentó una estructuración del dominio muy útil para abordar posteriormente el tema de la validez. Por razones de espacio comentaremos brevemente la segunda aproximación.

El trabajo de los jueces —expertos en el tema objeto del dominio— se estructuró en dos fases. La primera de ellas consistió en identificar las regiones o subdominios. Para este cometido aplicamos una estrategia de asignación basada en el índice de congruencia *ítem*-objetivo (Rovinelli y Hambleton, 1977). Valoraron, pues, la pertenencia de un objetivo a un subdominio. Una vez identificados los subdominios, se procedió a una doble valoración: interdominios e intradominio. En cada caso, se analizó la complejidad y relevancia de cada subdominio como unidad —primer caso— y entre los objetivos dentro de cada subdominio —segundo caso—. Para este cometido se utilizó un formato de juicio basado en la idea de «variables pivot» (Vicens Otero, 1981), de forma que se facilitaran las relaciones transitivas entre las variables. De esta manera, terminado el proceso de valoración, dispusimos de diversas matrices de juicio individual (una para cada subdominio y otra entre subdominios). Posteriormente, se extrajo para cada caso la matriz consenso, como la matriz de promedios de los jueces. Esta matriz presenta como características: cuadrada, simétrica, y se puede considerar como una matriz de similaridades.

Como análisis de dichas matrices realizamos:

a) En primer lugar, utilizamos Análisis de Escalamiento Multidimensional (MD-SCAL), dado que la matriz a estudiar es de juicio. Para este tipo de situaciones se ha demostrado (Carroll y Arabie, 1980; Carroll y Wish, 1982) que el MD-SCAL provee una información útil y precisa respecto a las asociaciones de las variables objeto de juicio sobre un plano formado por las dimensiones extraídas.

b) Escalamiento multidimensional no métrico (Kruskal, 1964 *a* y *b*, 1972; Kruskal y Wish, 1978; Guttman, 1954, 1968), dado que podíamos ga-

[13] Únicamente la complejidad cognitiva, como paragón de lo que podríamos denominar dificultad teórica, se podría analizar a partir del análisis de la dificultad empírica. Por su parte, la afinidad temática podría entenderse como nivel de rendimiento entre unidades y se podría expresar como correlaciones entre resultados empíricos, si se asume que la afinidad temática es una expresión de la generalidad del aprendizaje. Por su parte, relevancia o secuencialización parecen ser variables interpretables únicamente desde el análisis lógico.

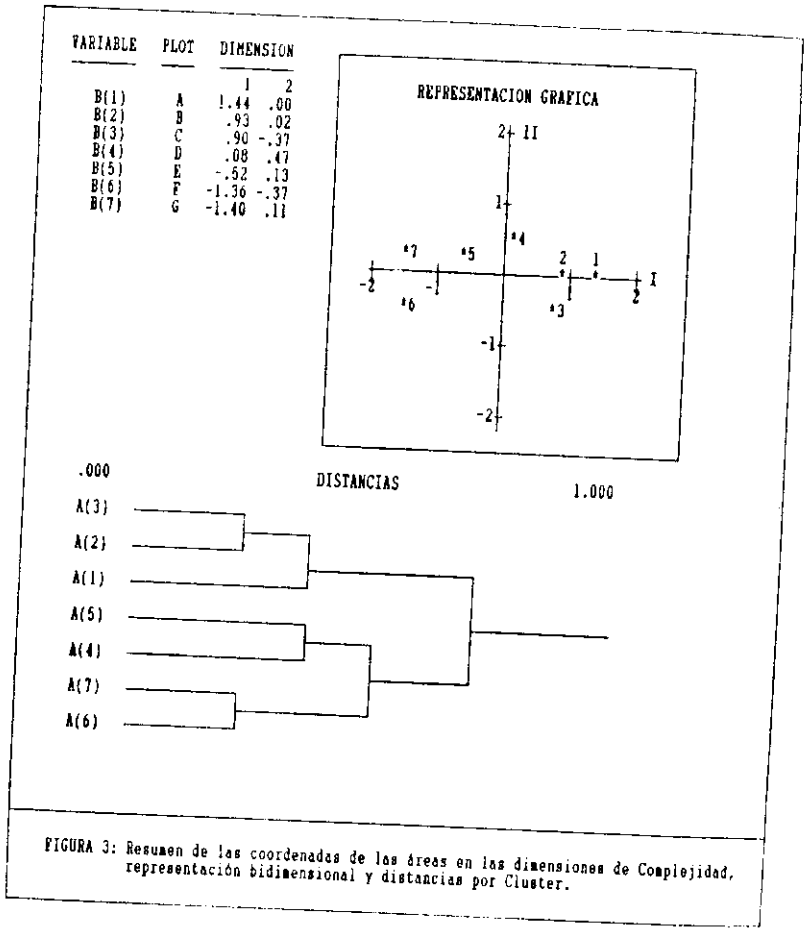


FIGURA 3: Resumen de las coordenadas de las áreas en las dimensiones de Complejidad, representación bidimensional y distancias por Cluster.

rantizar únicamente una escala ordinal para nuestras mediciones, pero no más allá (Green y Carmone, 1970) [14].

c) Como es sabido, existen dos métodos básicos alternativos dentro de este tipo de aproximación: 1. Método de regresión monótona de Kruskal (1964 *a*, 1964 *b*, 1972). 2. Método de permutación de Guttman (1954, 1968). De ellos, el más difundido es, sin duda alguna, el de Kruskal; Van der Ven (1980) los valora como equivalentes, aunque los métodos en sí mismos implican formas de cálculo diferentes. En este sentido, no parece haber evidencias que aconsejen la utilización de uno u otro método. Así, considerando que no teníamos experiencia previa en relación a la utilización de esta metodología con este tipo de objetos, que no conocemos otros trabajos que se hayan realizado en este ámbito y dado que, en general, tampoco se dan criterios respecto a las relaciones *stress*/número de dimensiones o alienación/número de dimensiones [15], optamos por seguir una estrategia analítica que nos permitiera seleccionar la estructura más representativa para cada variable.

En este caso, se siguieron las indicaciones de Spence (1982), que aconseja desarrollar un estudio de simulación con el fin de determinar los niveles mínimos de *stress* debidos al azar. Así, éstos pueden utilizarse, al menos, como un criterio de selección de la solución.

En nuestro caso encontramos soluciones más satisfactorias con el método de Kruskal. Por otra parte, una vez definidas las dimensiones obtenidas a partir de MD-SCAL, se estudiaron las asociaciones interdimensionales a partir de análisis cluster de variables [16]. De este modo, cada matriz consenso de juicio se analizó mediante MD-SCAL —método de Kruskal— y sus dimensiones a través de análisis cluster —ver figura 3—. Esta aproximación es útil cuando se pretende utilizar como una evidencia —en el sentido señalado por Hambleton (1984)— de validez de constructo, comparando los resultados del análisis de la estructura implícita con la resultante, dado que puede aplicarse sobre una matriz de distancias elaborada a partir de la de intercorrelaciones de los *items*, obtenido de la matriz vaciado de *items* original (sujetos-*items)* (Jornet, 1987).

En otro orden de cosas, en relación al análisis de la estructura resultante, se realiza a partir del análisis de los resultados empíricos del *test*. A tal efecto se han efectuado un mayor número de propuestas, si bien la mayoría de ellas giran en torno al análisis factorial para determinar la dimensionalidad del dominio. Estas van desde aplicaciones directas (Green, Lissitz y Mulaik, 1977; McDonald, 1981; Smith, 1980; Hambleton, 1984) hasta jerárquicas (Gorsuch y Dreger, 1979; Gorsuch y Yagel, 1982; Gorsuch, 1983), tanto con supuestos de ortogonalidad como de dependencia de los factores. En este punto, nos gustaría hacer alguna matización. Como señala López Feal (1986, pág. 342): «*... Un problema que se plantea cuando se utiliza el modelo de análisis factorial exploratorio o confirmatorio de*

[14] Téngase en cuenta que el análisis con MD-SCAL implica la transformación previa a la matriz de distancias, lo cual realizamos con distancias euclídeas utilizando el módulo CORR-SYSTAT. Se entienden como matrices de disconformidades.

[15] Existen diversas tablas para valorar la representatividad del valor de *stress* en relación al número de dimensiones, pero éstas son limitadas al número de objetos (Klahr, 1969; Stenson y Knoll, 1969; Wagenaar y Padmos, 1971, y Spence y Ogilvie, 1973).

[16] Téngase en cuenta que es difícil imaginar la distancias entre las variables en un espacio n-dimensional, y para este cometido, el análisis Cluster es un instrumento suplementario de gran utilidad (Dillon y Goldstein, 1984).

items *de un* test *unifactorial o multifactorial es el de la influencia de la dificultad del* item *sobre la correlación* item-item *o* item-test». Así, diversos autores analizan los problemas que implican dicha utilización (Ferguson, 1941; Wherry y Gaylord, 1944). En relación a estos problemas, Suárez (1987) los revisa y demuestra empíricamente; señala que éstos pueden resumirse en los siguientes puntos:

a) Se dará menor correlación entre dos *items* cuando haya mayor diferencia entre sus dificultades.
b) Así, el rango de la matriz de correlaciones depende de las diferencias de dificultad.

La consecuencia de estos puntos es que la matriz factorial resultante refleja las consistencias de dificultades, pero no las de contenidos. Ello, como es obvio, limita las interpretaciones de resultados empíricos a interpretaciones respecto a la dificultad y, generalizando el concepto, al nivel de complejidad de la temática. Así, a este respecto estimamos que el análisis cluster [17] podría ser más adecuado para este propósito y, en cualquier caso, se ajusta más a la lógica y necesidades del problema que estamos tratando.

Finalmente, hay que mencionar la aplicación de acercamientos basados en la aplicación del modelo de ecuaciones estructurales (Muthen, 1978, 1981, 1982; Muthen y Christofferson, 1981), que, como señala Roid (1984), están ofreciendo resultados alentadores.

4. Sistemas gestores

Dentro de lo que hemos denominado sistemas gestores, el primero en funcionar como tal, que fue desarrollado específicamente en el ámbito de los TRCs, es el Instructional Objetive Exchange (IOX) [18], creado por Popham en 1968. Fundado en la Universidad de California, es un centro especializado en el desarrollo de TRCs. En el mismo se ha creado, con el cúmulo de trabajo desarrollado, un banco de objetivos aplicables a diversas áreas y materias [19], con el fin de ofrecer a los diversos profesionales de la enseñanza los medios instrumentales para generar con garantías TRCs, derivando *items* de los objetivos [20] pertenecientes al dominio acotado.

Inicialmente, al percatarse de los problemas que se daban al tratar de generar *items* [21] a partir de objetivos operativos, intentaron aplicar metodologías que paliaran estos problemas al centrar mejor el dominio. Utilizaron para ello la metodología de Hively (formatos de *items*), así como la aproximación de Millman encaminada hacia la automatización, que se basaba tanto en formatos de *items* como en diseño de facetas. Sin embargo, según Popham (1980, 1984), tales metodologías guiaban la generación hacia áreas sumamente específicas. Así, planteó un tratamiento

[17] Como el propuesto por Guttman (1968), Lingoes (1977) o bien Berk (1978), aunque de forma específica dentro de la aproximación del diseño de facetas.
[18] Centro de intercambio de objetivos instruccionales (IOX).
[19] Vienen catalogadas tanto la enseñanza primaria como la secundaria, en todas sus materias.
[20] Cada objetivo incluye siempre ejemplos de *items*, alrededor de seis.
[21] Habían muchas diferencias entre los escritores de *items*.

intermedio entre los objetivos operativos y las metodologías antedichas, los cuales denominó objetivos amplificados, que en los estudios realizados han demostrado, sin embargo, tener una definición excesivamente ambigua. La opción final, hasta el momento, ha consistido en elaborar especificaciones de la prueba de manera que sean suficientemente concretas como para comunicarse adecuadamente con los constructores de la misma, y suficientemente ajustadas a la realidad (dominio suficientemente amplio) como para que sean útiles en la mayoría de las situaciones que pueden darse en una clase.

Paralelamente al trabajo de Popham con el IOX, se han venido desarrollando diversas opciones de gestión de dominios. En este sentido, identificamos como sistemas gestores: *a)* bancos de *items*, *b)* programas generadores de *items* y *c) tests* administrados y/o asistidos por ordenador (CAT).

En estas áreas, Millman (1974, 1980, 1984) realiza una amplia evaluación y revisión de sus requerimientos y utilidades para los TRC.

En primer lugar, respecto a los bancos de *items*, si bien su desarrollo ha sido independiente de los TRC, se viene introduciendo, aunque lentamente, en este ámbito (Millman, 1974, 1982; Millman y Outlaw, 1978; Millman y Arter, 1984; Finn, 1975; Hsu y Nitko, 1983; Nitko y Hsu, 1984; Rodríguez Lajo, 1986). No vamos a realizar aquí una revisión de los métodos propios para definir y gestionar bancos de *items*, dado que ello constituye en sí mismo un área de investigación que si bien es útil para la ERC, su desarrollo y tecnología son independientes de ésta. Es más, la variedad de sistemas es amplia y existen concepciones diversas respecto a cómo construir bancos de *items* computerizados que oscilan desde diseños más rígidos, como el de Millman (1980, 1984) y Millman y Arter (1984), a otros más flexibles con una concepción interactiva, como el propuesto por Nitko y Hsu (1984). En este punto únicamente queríamos resaltar las aproximaciones de los bancos de *items* como posibles sistemas gestores de dominios educativos bien definidos y, en ese sentido, como una tecnología que debería considerarse seriamente en su aplicación en la ERC.

Como una evolución del tema anterior, pueden contemplarse algunas tecnologías de las revisadas con anterioridad, como, por ejemplo, los formatos de *items* o el diseño de facetas, los cuales permiten, sin escribir directamente los *items*, poseer un banco de procedimientos generadores cuya ventaja más relevante es que pueden formularse *items* de contenido similar con dificultad equivalente —ver cuadro 3—. El sistema gestor que actúe como soporte para este tipo de procedimiento puede ser del tipo tradicional o bien computerizado. Este último supone escribir programas que generan los *items* (Millman, 1974, 1980, 1984) y presenta como ventaja, respecto a los bancos de *items*, precisamente el hecho de que no es precisa la definición pormenorizada de todas las unidades de elementos. Evidentemente, en un banco de *items* se asocia a cada uno de ellos información métrica (p_i, σ_i^2, etc.), que va a ser utilizada para la selección de elementos cuando se desean construir pruebas paralelas clásicas. A este tipo de acercamiento se le ajusta mejor la asunción relajada de pruebas aleatoriamente paralelas, pues asume que todos los elementos generables a partir de un modelo son equivalentes. Una ventaja adicional de los programas de generación de *items* sobre los bancos de *items* reside en que al aumentar el número de elementos generables, también aumenta, en consecuencia, el número de pruebas paralelas que pueden desarrollarse

a partir de ellos. Por otra parte, es preciso considerar que también sería susceptible la programación de elementos métricos adicionales que sirvieran para la selección de *ítems* (en este caso, modelos) —ver figura 4—, los cuales necesariamente se situarían a un nivel jerárquico superior dentro de la estructura del paquete gestor.

Finalmente, los *tests* administrados y/o asistidos por computador (CAT) pueden entenderse como el extremo en la evolución de gestión computerizada de dominios. Así, hay que considerar que junto a los *ítems* —o los programas generadores—, deben incluir los elementos necesarios para que el programa pueda ir tomando decisiones de presentación a partir de la evaluación. El nivel de sofisticación, pues, es altísimo y, desde nuestro punto de vista, sólo sería aplicable en temas educativos con diseños en los que se tuviera una gran certidumbre respecto a: *a)* definición del dominio, *b)* definición de las estrategias cognitivas subyacentes al mismo, *c)* estructura del dominio, y *d)* estructura secuencial, descrita en los términos de habilidad y de estrategia, y, por otra parte, respecto a la prueba: *a)* criterios de bondad, *b)* estándares definidos en la escala de puntajes verdaderos, y *c)* criterios de corte [22] basados en la teoría de la decisión, con definición del tipo de pérdida [23] que se asume, valorando su ade-

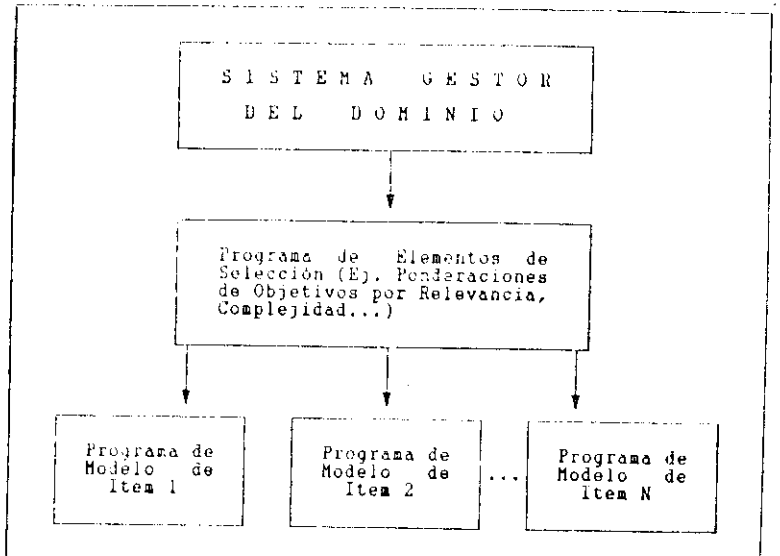


FIGURA 10: Jerarquía de programas para Sistemas Gestores de Dominios a partir de Programas Generadores de Items (Paquete Gestor).

[22] Para este cometido, Millman (1984) reseña el procedimiento de Wald (1947), el cual está extraído de un procedimiento para determinar el análisis secuencial de productos dentro de un control de calidad.

[23] Asociada a la consideración del error, como pérdida en umbral, de error al cuadrado, etc.

cuación al dominio. Bajo estas precauciones estimamos que pueden desarrollarse los CAT en la ERC.

Con independencia de estas precauciones queríamos señalar que el trabajo realizado al respecto (Lord, 1971; Weiss y Betz, 1973; Weis y Kingsbury, 1984; Cliff, 1975; Cudeck, Cliff y Kehoe, 1977; McCormick y Cliff, 1977; Baker, 1984; McArthur y Choppin, 1984, entre otros) es prometedor y, sin duda alguna, presenta realizaciones de gran utilidad.

BIBLIOGRAFIA

Anderson, R. C. (1972). How to construct achievement tests to assess comprehension. *Review of Educational Research*, 42, 145-170.

Baker, F. B. (1984). Technology and Testing: State of the Art and Trends for the Future. *Journal of Educational Measurement*, 21 (4), 399-406.

Béjar, I. (1983). Achievement Testing, Recent Advances. *Sage University Papers.* Serie: *Quantitative Applications in the Social Sciences.*

Berk, R. A. (1978). The application of structural facet theory to achievement test construction. *Educational Research Quarterly*, 3, 62-72.

Berk, R. A. (1980). A comparison of six content domain specification strategies for criterion-referenced Tests. *Educational Technology*, 20, 49-52.

Besel, R. (1973). *Using group performance to interpret individual responses to criterion-referenced tests.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Besel, R. (1975). *Mixed group validation and the problem of mastery-learning decisions.* Paper presented at the annual meeting of the American Educational Research Association, Washington, DC.

Birenbaum, M., y Tatsuoka, K. (1982). On the dimensionality of achievement test data. *Journal of Educational Measurement*, 19, 259-266.

Birenbaum, M., y Tatsuoka, K. (1983). The effect of a scoring system based on the algoritm underlying the students' response patterns on the dimensionality of achievement test data of the problem solving type. *Journal of Educational Measurement*, 20, 17-26.

Block, J. H. (Ed.) (1971). *Mastery Learning: Theory and Practice.* New York: Holt. Rinehart and Winston.

Bloom, B. S. (1968). Learning for mastery. *Evaluation Comment*, 1, 1-12.

Bloom, B. S. (1971). Mastery Learning. En J. H. Block (Ed.), *Mastery Learning: Theory and Practice.* New York: Holt. Rinehart y Winston.

Bloom, B. S. (1974). Time and learning. *American Psychologist*, 29, 682-688.

Bloom, B. S., *et al.* (1969). *Taxonomie des objectifs pedagogiques: le domaine cognitif.* Montreal: Education Nouvelle.

Bloom, B. S.; Engelhart, M. D.; Furst, E. J.; Hill, W. H., y Krathwohl, D. R. (1956). *Taxonomy of Educational Objectives: the Cognitive Domain.* New York: Longmans, Green.

Bloom, B. S.; Hastings, J. T., y Madaus, G. (1971). *Handbook of Formative and Summative Evaluation of Student Learning.* New York: McGraw-Hill.

Bloom, B. S.; Madaus, G. F., y Hastings, J. T. (1981). *Evaluation to improve Learning.* New York: McGraw-Hill.

Bormuth, J. R. (1970). *On the theory of achievement Test Items.* Chicago, Illinois: Univ. of Chicago Press.

Braby, R.; Parrish, W. F.; Guitard, C. R., y Aagard, J. A. (1978). *Computer-aided authoring of programmed instruction for teaching symbol Recognition* (Tech. Report N.º 58). Orlando, Florida: Training Analysis and Evaluation.

Bradfield, J. H., y Moredock, H. S. (1957). *Measurement and Evaluation in Education and Introduction to its Theory and Practice at both the Elementary and*

Secundary Schools levels. MsMillan, New York.
Brown, S. S., y Burton, R. R. (1978). Diagnostic models procedural bugs in basic mathematics Skills. *Cognitive Science*, 2, 155-192.
Carroll, J. B. (1963). A model for school learning. *Teachers College Record*, 64, 723-733.
Carroll, J. B. (1976). Psychometric tesis as cognitive tasks: A new «structure of intellect». En L. B. Resnick (Ed.), *The nature of intelligence*. New York: Jhon Wiley, 27-56.
Carroll, J. B. (1981). Ability and task difficulty in cognitive psychology. *Educational Research*, 10, 11-21.
Carroll, J. D., y Arabie, P. (1980). Multidimensional Scaling. En M. R. Rosenzweig y L. W. Porter (Eds.), *Annual Review of Psychology*, 31, 607-649.
Carroll, J. D., y Wish, M. (1982). Multidimensional perceptual models and measurement methods. En P. M. Davies y A. P. M. Coxon (Eds.), *Key Texts in Multidimensional Scaling*. London: Heinemann Educational Books Ltd., 43-58.
Cliff, N. (1975). Complete orders from incomplete data: Interactive ordering and tailored testing. *Psychological Bulletin*, 82, 289-302.
Coffman, W. E. (1971). Essay examinations. En R. L. Thorndike (Ed.), *Educational measurement* (2.ª). Washington, DC: American Council on Education, 271-302.
Conoley, J. C., y O'Neil, H. F., Jr. (1979). A primer for developing test items. En H. F. O'Neil, Jr. (Ed.), *Procedures for instructional systems development*. New York: Academic Press, 95-127.
Cronbach, L. J. (1970). Review of on the theory of achievement tests items. En J. R. Bormuyh (Ed.). *Psychometrika*, 35, 509-511.
Cudeck, R. A.; Cliff, N., y Kenoe, J. F. (1977). TAILORR: A fortran procedure for interactive tailored testing. *Educational and Psychological Measurement*, 37, 767-769.
Cudeck, R. A.; McCormick, D. J., y Cliff, N. (1980). Implied Orders Tailored Testing: Simulation with the Stanford-Binet. *Applied Psychological Measurement*, 4 (2), 157-163.
De la Orden Hoz, A. (1983). La investigación sobre la evaluacion educativa». *Revista de Investigación Educativa*, 1 (2), 240-258.
Dillon, W. R., y Goldstein, M. (1984). *Multivariate Analysis: Methods and Applications*. New York: Wiley.
Drum, P. A.; Calfee, R. C., y Cook, L. K. (1980). The effects of surface structure variables on performance in reading comprehension test. *Reading Research Quarterly*, 16, 486-513.
Durnin, J. H., y Scandura, J. M. (1973). An algorithm approach to assessing behavior potential: Comparison with item forms and hierarchical technologies. *Journal of Educational Psychology*, 64, 262-272.
Ebel, R. L. (1979). *Essentials of Educational Measurement* (3.ª ed.). Englewood Cliffs, N.J.: Prentice-Hall.
Ellis, J. A., y Wulfeck, II. W. J. (1978). *The Instructional Quality Inventory: IV. Job performance AID* (NPRCD Special Report 79-5). San Diego: Navy Personnel Research and Development Center. [Available from Defense Technical Information Center (DTIC) document number AD A083928.]
Ellis, J. A., y Wulfeck, W. H. (1980). *Assuring Objective-Test Consistency: A Systematic Procedure for Constructing Criterion-Referenced Tests*. San Diego, California: NPRDC. Special Report, 80-15.
Ellis, J. A., y Wulfeck, W. H. (1982). *Handbook for testing Navy Schools*. San Diego, CA: Navy Personnel Research and Development Center.
Ellis, J. A.; Wulfeck, W. H., y Fredericks, P. S. (1979). *The Instructional Quality Inventory: Vol. II, User's Manual*. San Diego: NPRDC, Special Report, 79-24. (DTIC AD A085678.)
Embretson, S. E. (Ed.) (1985). *Test Design: Development in Psychology and Psychometrics*. Florida: Academic Press, Inc.
Emrick, J. A. (1971). An evaluation model for mastery testing. *Journal of Educational Measurement*, 4, 321-325.

Emrick, J. A., y Adams, E. N. (1969). *An evaluation model for individualized instruction* (Report RC 2674). Yorktown Hts., NY: IBM, Thomas J. Watson Research Center.
Engel, J. D., y Martuza, V. R. (1976). *A systematic approach to the construction of Domain-Referenced Multiple-Choice test items*. Paper presented at the meeting of the American Psychological Association, Washington, DC, Spr.
Evans, J. St. B. T. (1982). *The Psychology of Deductive Reasoning*. London: Routledge and Kegan Paul.
Ferguson, G. A. (1941). The factorial interpretation of test difficulty. *Psychometrika*, 6, 323-329.
Finn, P. J. (1975). A question writing algoritm. *Journal of Reading Behavior*, 341-367.
Fleishman, E. A., y Quaintance, M. K. (1984). *Taxonomies of human performance (The description of Human tasks)*. Academic Press, Inc., London.
Frederiksen, C. H. (1975). Representing logical and semantic structure of knowledge adquired from discourse. *Cognitive Psychology*, 7, 371-458.
Fremer, J., y Anastasio, E. J. (1969). Computer-assisted item writing-I (Spelling items). *Journal of Educational Measurement*, 6, 69-74.
Gagne, R. M. (1971). *Defining Objectives for six types of Learning*. Washington, DC: American Educational Research Association.
García Hoz, V. (1970). *Principios de Pedagogía Sistemática*. Madrid: Rialp.
Gigk, M. L., y Holyoaka, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, 12, 306-355.
Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist*, 18, 519-521.
Glaser, R. (1976). Components of a Psychology of Instruction: Toward a Science of Design. *Review of Educational Research*, 46 (1), 1-24.
Glaser, R. (1981). A research agenda for cognitive psychology and psychometrics. *American Psychologist*, 36, 923-936.
Gorsuch, R. L. (1983). *Factor analysis* (2.ª ed.). Hillsdale, NJ: Erlbaum.
Gorsuch, R. L., y Dreger, R. M. (1979). «Big jiffy: A more spphisticated factor analysis and rotation program». *Educational and Psychological Measurement*, 39, 209-214.
Gorsuch, R. L., y Yagel, J. C. (1982). *Exploratory item factor analysis*. Unpublished paper, Fuller Graduate School of Psychology, Pasadena, California.
Green, P. E., y Carmone, F. J. (1970). *Multidimensional Scaling and related techniques in marketing analysis*. Boston, Allyn and Bacon.
Green, S. B.; Lissitz, R. W., y Mulaik, S. A. (1977). Limitation of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement*, 37, 827-838.
Gronlund, N. E. (1982). *Constructing achievement tests* (3.ª ed.) Englewood Cliffs, NJ: Prentice-Hall.
Guilford, J. P. (1956). The structure of intellect. *Psychological Bulletin*, 53, 267-293.
Guilford, J. P. (1967). *The Nature of Human Intelligence*. New York: McGraw-Hill.
Guilford, J. P. (1981). Higher-order structure of intellect abilities. *Multivariate Behavioral Research*, 16, 411-435.
Guskey, T. R. (1980 *a*). Mastery learning: Applying the theory. *Theory into Practice*, 19, 104-111.
Guskey, T. R. (1980 *b*). What is mastery learning. *Introductor*, 90 (3), 80-84.
Guskey, T. R. (1982). The theory and practice of mastery learning. *The principal*, 27 (4), 1-12.
Guskey, T. R. (1985). *Implementing mastery learning*. Wadsworth, P.C. Belmony, California.
Guttman, L. (1954). A new approach to factor analysis: The radex. En P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences*. New York: Free Press.

Guttman, L. (1959). A structural theory for intergroup beliefs and actions. *American Sociological Review*, 24, 318-328.
Guttman, L. (1965). *The structure of interrelations among intelligence tests.* Proceedings of the 1964 Invitational Conference on Testing Problems. Princeton, New Jersey: Educational Testing Service.
Guttman, L. (1968). A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika*, 33, 469-506.
Guttman, L. (1969). *Integration of test design and analysis.* Proceedings of the 1969 Invitational Conference on Testing Problems. Princeton, New Jersey: Educational Testing Service.
Guttman, L. (1970). *Integration of test design and analysis.* Proceedings of the 1969 Invitational Conference on Testing Problems. Princeton, New Jersey: Educational Testing Service.
Hambleton, R. K. (1984). Criterion-referenced measurement. En T. Husen y T. N. Postlethwaite (Eds.), *International Encyclopedia of Education: Research and studies.* Oxford, England: Pergamon Press.
Hambleton, R. K. (1986). Criterion-referenced assessment of individual differences. En C. R. Reynolds y V. L. Willson (Eds.), *Methodological and Statistical advances in the study of individual differences.* Plenum Press, New York.
Hambleton, R. K., y Novick, M. R. (1973). Toward an integration of theory and method for criterion-referenced tests». *Journal of Educational Measurement*, 10, 159-170.
Hively, W. (1966). *A Test-Item pool for Minnemast Science UNIT 2.1: Measuring weight.* Unpublished paper. Minnemast Project. Univ. of Minnesota.
Hively, W.; Patterson, H. L., y Page, S. A. (1968). A «Universedefined» system of aritmetic achievement tests. *Journal of Educational Measurement*, 5, 275-290.
Hoepfner, R. (1978). Achievement test selection for program evaluation. En M. J. Wargo y D. R. Green (Eds.), *Achievement testing of disadvantaged and minority students for educational program evaluation.* Monterrey, CA: CTB/McGraw-Hill.
Hsu, T. C., y Nitko, A. J. (1983). *Microcomputer testing software teachers can use.* Paper presented at the ECS Large-scale Assessment Conference, Boulder, Co.
Huynh, H. (1976). Statistical consideration of mastery scores. *Psychometrika*, 41, 65-78.
Huynh, H. (1980). A nonrandomized minimax solution for passing scores in the binomial erro model. *Psychometrika*, 45 (2), 167-182.
Johnson, K. J. (1973). Pitt's computer-generated chemistry exam. *Proceedings of the Conference on Computers in Undergraduate Curricula*, 199-204.
Johnson-Laird, P. N. (1983). *Mental Models.* Cambridge: Cambridge University Press.
Johnson-Laird, P. N., y Steedman, M. (1978). The psychology of syllogisms. *Cognitive Psychology*, 10, 64-98.
Jornet, J. M. (1980). *Los tests psicométricos como indicadores cognitivos.* Tesis de licenciatura no publicada. Universidad de Valencia.
Jornet, J. M. (1987). *Una aproximación teórico-empírica a los métodos de medición de referencia criterial.* Tesis doctoral no publicada. Universidad de Valencia.
Jornet, J. M., y Suárez, J. M. (1984). Una caracterización cognitiva por niveles de rendimiento de tests psicométricos. *Millars*, IX (1-2), 43-74.
Klahr, D. (1969). Monte Carlo investigation of the statistical significance of Kruskal's nonmetric multidimensional scaling algorithms. *Psychometrika*, 37, 461-486.
Kruskal, J. B. (1964 *a*). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29, 1-27.
Kruskal, J. B. (1964 *b*). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 29, 115-129.
Kruskal, J. B. (1972). Linear transformation of multivariate data to reveal clustering. En R. N. Shepard, A. K. Romney y S. B. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences.* Vol. I. New York: Seminar Press.

Kruskal, J. B., y Wish, M. (1978). *Multidimensional scaling.* Murray Hill, New Jersey: Bell Laboratories.
Lewis, C.; Wang, M. M., y Novick, M. R. (1973). Marginal distributions for the estimation of proportions in groups. *ACT Technical Bulletin.* No. 13. Iowa City, Iowa: The American College Testing.
Lingoes, G. C. (1977). *Geometric representations of relational data.* Ann Arbor, MI: Mathesis Press.
Linn, R. L. (1983). Curriculum validity: Convincing the court that it was taught without precluding the possibility of measuring it. En G. F. Madaus (Ed.), *The courts, validity and minimum competency testing.* Hingham, MA: Kluwer-Hijhoff, 115-132.
López Feal, R. (1986). *Construcción de instrumentos de medida en ciencias conductuales y sociales.* Vol. I. Barcelona: Alamex.
Lord, F. M. (1971). The self-scoring flexilevel test. *Journal of Educational Measurement*, 8, 147-151.
Macready, G. B., y Dayton, C. M. (1977). The use of probabilistic models in the assessment of mastery. *Journal of Educational Statistics*, 2, 99-120.
Macready, G. B., y Dayton, C. M. (1980 *a*). The Nature and Use of State Mastery Models. *Applied Psychological Measurement*, 4 (4), 493-516.
Macready, G. B., y Dayton, C. M. (1980 *b*). A two-stage conditional estimation procedure for unrestricted latent class models. *Journal of Educational Statistics*, 33, 351-360.
Madaus, G. F. (Ed.) (1983). *The courts, validity, and minimum competency testing.* Higham, MA: Kluwer-Nijhoff.
Mager, R. F. (1973). *Measuring Instructional intent or. Got a Match?* Belmont, California: Lear Siegler, Inc./Fearon Publishers.
Markle, S. M., y Tiemann, P. W. (1970). *Really Understanding concepts.* Champaign, Illinois: Stipes Publishing Company.
Martuza, V. (1979). *Domain definition/item generation for criterion-referenced tests: A review and directions for future resarch.* Paper presented at the annual meeting of the Eastern Educational Research Association. Kiawah Island, South Carolina, February.
McArthur, D. L., y Choppin, B. H. (1984). Computerized Diagnostic Testing. *Journal of Educational Measurement*, 21 (4), 391-398.
McClain, D. H.; Wessels, S. W., y Sando, K. M. (1975). *IPSIM-Additional system enhancements utilized in a chemistry application.* Procedings of the Conference on Computers in Undergraduate Curricula, 139-145.
McClung, M. S. (1978). *Developing proficiency programs in California public schools: Some legal implications and a suggested implementation schedule.* Sacramento, CA: California Department of Education.
McCormick, D. J., y Cliff, N. (1977). TAILOR-APL: An interactive computer program for individual tailored testing. *Educational and Psychological Measurement*, 37, 771-774.
McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology*, 34, 100-117.
Meeker, M., y Meeker, R. (1975). *Structure of Intellect (SOI) Learning Abilities Tests.* El segundo, CA: SOI Institute.
Mellenbergh, G. J.; Kopelar, H., y Van der Linden, W. J. (1977). Dichotomous decisions based on dichotomously scored items: A case study. *Sttatistica Neerlandica*, 31, 161-169.
Mellenbergh, G. J., y Van der Linden, W. J. (1981). The linear utility model for optimal selection. *Pychometrika*, 46 (3), 283-293.
Merrill, M. D.; Reigeluth, C. M., y Faust, G. W. (1979). The instructional quality profile: A curriculum evaluation and design tool. En H. F. O'Neil, Jr. (Ed.), *Procedures for Instructional Systems Development.* New York: Academic Press.
Miller, W. G.; Snowman, J., y O'Hara, T. (1979). Application of alternative statistical techniques to examine the hierarchical ordering in Bloom's taxonomy. *American Educational Research Journal*, 16, 241-248.

Miller, H. G., y Williams, R. G. (1973). Constructing higher level multiple choice questions covering factual content». *Educational Technology*, 13 (5), 39-42.
Miller, F. G.; Williams, R. G., y Haladyna, T. M. (1978). *Beyond FACTS: Objective ways to measure thinking*. Englewood Cliffs, New Jersey: Educational Technology Publications.
Millman, J. (1974). Criterion-referenced measurement. En W. J. Popham (Ed.), *Evaluation: Current Applications*. Berkeley, California: McCutchan Publishing Company.
Millman, J. (1980). Computer-based item generation. En R. A. Berk (Ed.), *Criterion-referenced Measurement*. Baltimore, Maryland: Johns Hopkins, University Press.
Millman, J. (1982). A system for generating unique tests by computer and its use in a mastery learning setting. En P. R. Baumann (Ed.), *Computer Based Instruction*. Albany: Faculty Grants for the Improvement of Undergraduate Instruction, State University of New York, 99-106.
Millman, J. (1984). Individualizing Test Construction and administration by computer. En R. A. Berk (Ed.), *A guide to criterion-referenced test construction*. Baltimore: The Johns Hopkins. University Press.
Millman, J., y Arter, J. A. (1984). Issues in Item Banking. *Journal of Educational Measurement*, 21 (4), 315-330.
Millman, J., y Outlaw, W. S. (1978). Testing by computer. *Association for educational data systems (AEDS) Journal*, 11, 57-72.
Montague, W. E.; Ellis, J. A., y Wulfeck, W. H. (1983). Instructional quality inventory: A formative evaluation tool for instructional development. *Performance and Instruction*, 22 (5), 11-14.
Muthen, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika*, 43, 551-560.
Muthen, B. (1981). Factor analysis of dichotomous variables: American attitudes toward abortion. En E. Borgatta y D. J. Jackson (Eds.), *Factor Analysis and measurement in Sociological Research: A multidimensional perspective*. San Francisco: Sage.
Muthen, B. (1982). *LACCI: Latent variable analysis with dichotomous, ordered categorical and continous indicators*. Los Angeles: Graduate School of Education. University of California.
Muthen, B., y Christofferson, A. (1981). Simultaneous factor analysis of dichotomous variables in several groups. *Psychometrika*, 46, 407-419.
Newell, A. (1980). Reasoning, problem solving, and decision processes: The problem space as a fundamental category. En R. Nickerson (Ed.), *Attention and Performance*. Vol. VIII. Hillsdale, New Jersey: Lawrencen Erlbaum Associates.
Nitko, A. J. (1980). Distinguishing the many varieties of criterion-referenced tests. *Review of Educational Research*, 50, 461-485.
Nitko, A. J. (1984). Defining criterion-referenced test. En R. A. Berk (Ed.), *A guide to criterion-referenced test construction*. Baltimore: The Johns Hopkins, University Press, 8-28.
Nitko, A. J., y Hsu, T. (1984). A comprehensive Microcomputer System for Classroom Testing. *Journal of Educational Measurement*, 21 (4), 377-390.
Novick, M. R.; Lewis, Ch., y Jackson, P. H. (1973). The estimation of proportions in groups. *Psychometrika*, 38 (1), 19-46.
Novick, M. R., y Lindley, D. V. (1978). The use of more realistic utility functions in educational applications. *Journal of Educational Measurement*, 15, 181-191.
Novick, M. R., y Jackson, P. H. (1974). *Statistical Methods for Educational and Psychological Research*. New York: McGraw-Hill.
Olympia, P. L., Jr. (1975). Computer generation of truly repeteable examinations. *Educational Technology*, 14 (6), 53-55.
Osburn, H. G. (1968). Item sampling for achievement testing. *Educational and Psychological Measurement*, 28, 95-104.
Popham, W. J. (1974). Selecting objetives and generating test intems for objectives-based tests. En *Problems in criterion-referenced Measurement ed. C.* W. Harris, M. C. y W. J. Popham. CSE Monograph Series in Evaluation, n.º 3, Los Angeles: Center for the study of Evaluation, University of California, 13-25.
Popham, W. J. (1975). *Educational Evaluation*. Englewood Cliffs, New Jersey: Prentice Hall.
Popham, W. J. (1978). *Criterion-referenced Measurement*. Englewood Cliffs, New Jersey: Prentice Hall. (Traducción castellana. *Evaluación basada en criterios*. Ed. Magisterio Español, S. A., 1983, Madrid.)
Popham, W. J. (1980). Content domain specifications. En R. A. Berk (Ed.), *Criterion-referenced measurement: The state of the art*. Baltimore, MD: Johns Hopkins, University Press, 15-31.
Popham, W. J. (1984). Specifying the domain of content or behaviors. En R. A. Berk (Ed.), *A guide to Criterion-referenced test construction*. Baltimore, MD: Johns Hopkins, University Press, 29-48.
Popham, W. J., y Baker, E. L. (1970). *Systematic Instruction*. Englewood Cliffs, New Jersey: Prentice Hall.
Popham, W. J., y Husek, T. R. (1969). Implications of criterion-referenced measurement. *Journal of Educational Measurement*, 6 (1), 1-9.
Rips, L. J., y Markus, S. L. (1977). Suppositions and the analysis of conditional sentences. En M. A. Just y P. A. Carpenter (Eds.), *Cognitive Processes in Comprehension*. New York: Wiley.
Rodríguez Diéguez, J. L. (1979). *Cuadernos de didáctica I: Objetivos educativos*. Valencia: Instituto de Ciencias de la Educación.
Rodríguez Lajo, M. (1986). *Incidencia de la evaluación en el rendimiento de los universitarios en estadística aplicada*. Resumen de la tesis doctoral. Barcelona: Publicacions Edicions Universitat de Barcelona.
Roid, G. H. (1984). Generating the test items. En R. A. Berk (Ed.), *A guide to criterion referenced test construction*. Baltimore: The Johns Hopkins, University Press, 49-77.
Roid, G. H., y Haladyna, T. M. (1978). A comparison of objective-based and modified-Bormuth item writing techniques. *Educational and Psychological Measurement*, 35, 19-28.
Roid, G. H., y Haladyna, T. M. (1979). *Handbook on item writing for criterion-referenced testing*. San Diego, California: Navy Personnel Research and Development Center.
Roid, G. H., y Haladyna, T. M. (1980 *a*). *Handbook of item writing for criterion-referenced tests*. San Diego, California: NPRDC, Technical Note, 80-8.
Roid, G. H., y Haladyna, T. M. (1980 *b*). The emergence of a technology of tests item writing. *Review of Educational Research*, 50, 293-314.
Roid, G. H., y Haladyna, T. M. (1982). *A technology for test-item writing*. New York: Academic Press.
Roid, G. H.; Haladyna, T. M., y Shaughnessy, J. A. (1980). *A comparison of item-writing methods for criterion-referenced testing*. Paper presented at the annual meeting of the National Council on Measurement in Education. Boston, April.
Roid, G. H.; Haladyna, T. M.; Shaughnessy, J., y Finn, P. (1979). *Item ovriting for domain-referenced tests of prose learning*. Paper presented at the annual meeting of the American Educational Research Association. San Francisco.
Roudabusch, G. E. (1974). *Models for a beginning theory of criterion-referenced tests*. Paper presented at the annual meeting of the National Council on Measurement in Education. Chicago.
Rovinelli, R. J., y Hambleton, R. K. (1977). On the use of content specialists in the assessment of criterion-refrenced test item validity. *Dutch Journal of Educational Research*, 2, 49-60.
Runkel, P. J., y McGrath, J. E. (1972). *Research on Human Behavior: a Systematic Guide to Method*. New York: Holt.
Scandura, J. M. (1970). Role of rules in behavior: Toward an operational definition of what (rule) is learned. *Psychological Review*, 77, 516-533.

Scandura, J. M. (1977). *Problem-solving: a Structural Process approach with Educational Implications.* New York: American Press.
Scandura, J. M., y Durnin, J. (1978). Assessing behavior potential adequacy of basic theoretical assumptions. *Journal of Structural Learning,* 6, 3-47.
Schmidt, W. H.; Porter, A. C.; Schwille, J. R.; Floden, R. E., y Freeman, D. J. (1983). Validity as a variable: Can the same certification test be valid for all students? En G. F. Madaus (Ed.), *The courts, validity, and minimum competency testing.* Higham, MA: Kluwer-Nijhoff, 133-151.
Schulz, R. E. (1979). Computer aids for developing tests and instruction. En H. F. O'Neil, Jr., *Procedures for Instructional Systems Development.* New York: Academic Press.
Scriven, M. (1967). The methodology of evaluation. En R. Tyler y cols. (Eds.), *Perspectives of curriculum evaluation.* Rand McNally. Chicago. Monograph Series of Curriculum Evaluation, 1.
Simon, H. A., y Reed, S. K. (1976). Modelling strategy shift in a problem-solving task. *Cognitive Psychology,* 8, 86-97.
Smith, J. K. (1980). On the examination of test unidimensionality. *Educational and Psychological Measurement,* 40, 885-889.
Spence, I. (1982). A simple aproximation of random rankings STRESS values. En Davies, P. M., y Coxon, A. P. M. (Eds.), *Key texts in multidimensional scaling.*
Spence, I., y Ogilvie, J. C. (1973). A table of expected stress values for random rankings in Kruskal's nonmetric scaling procedure. *Psychological Bulletin,* 72, 122-126.
Stenson, H. H., y Knoll, R. L. (1969). Goodness of fit for random rankings in Kruskal's nonmetric scaling procedure. *Psychological Bulletin,* 72, 122-126.
Sternberg, R. J. (1977). *Intelligence, information processing and analogical reasoning: The componential analysis of human abilities.* Hillsdale, New Jersey: Erlbaum.
Sternberg, R. J. (1979). The nature of mental abilities. *American Psychologists,* 34, 214-230.
Sternberg, R. J. (1980). Sketch of a componential subtheory of human intelligence. *Behavioral and Brain Sciences,* 3, 573-584.
Sternberg, R. J. (1982). A componential approach to intelectual developement. En R. J. Sternbergh (Ed.), *Advances in the psychology of human intelligence.* Vol. 1. Hillsdale, New Jersey: Erlbaum, 413-463.
Sternberg, R. J. (1984). Toward a triarchic theory of human intelligence. *The behavioral and Brain Sciences,* 7, 269-315.
Suárez, J. M. (1987). *Estudio psicométrico-diferencial de la escala WISC-R.* Tesis doctoral, no publicada. Universidad de Valencia (por cortesía del autor).
Swaminathan, H.; Hambleton, R. K., y Algina, J. (1975). A bayesian decision-theoretic procedure for use with criterion-referenced tests. *Journal of Educational Measurement,* 12 (2), 87-98.
Tiemann, P. W., y Markle, S. M. (1978 *a*). *Analysing Instructional content: A guide to Instruction and Evaluation.* Champaign, Illinois: Stipes Publishing Company.
Tiemann, P. W., y Markle, S. M. (1978 *b*). *Domain-referenced testing of conceptual learning.* Paper presented at the annual meeting of the American Educational Research Association. March. Toronto.
Tiemann, P. W., y Markle, S. M. (1983). *Analysing instructional content: A guide to instruction and evaluation* (2.ª ed.). Champaign, Illinois: Stipes.
Tversky, A., y Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science,* 185, 1124-1131.
Tversky, A., y Kahneman, D. (1978). Causal schemata in judgments under uncertainty. En M. Fishbein (Ed.), *Progress in Social Psychology.* Hillsdale, New Jersey: Lawrence Erlbaum Associates Inc.
Van der Linden, W. J. (1980). Decision models for use with criterion-referenced tests. *Applied Psychological Measurement,* 4 (4), 469-492.

Van der Linden, W. J. (1984). Some thought on the use of decision theory to let cutoff scores: Comment on De Gruijter and Hambleton. *Applied Psychological Measurement,* 8 (1), 9-17.
Van der Linden, W. J., y Mellenbergh, G. J. (1977). Optimal cutting scores using a linear loss function. *Applied Psychological Measurement,* 1, 593-599.
Van der Ven, A. H. G. S. (1980). *Introduction to scaling.* Chichester: Wiley.
Vicens Otero, J. (1981). Análisis multidimensional no métrico. En E. Ortega Martínez (Ed.), *Manual de investigación comercial.* Madrid: Pirámide.
Vickers, F. D. (1973). Creative test generators. *Educational Technology,* 13 (3), 43-44.
Wagenaar, W. A., y Padmos, P. (1971). Quantitative interpretation of stress in kruskal's multidimensional scaling technique. *British Journal of Mathematical and Statistical Psychology,* 24, 101-110.
Wald, A. (1947). *Sequential analysis.* New York: Wiley.
Wasik, J. L. (1979). GENTEST: A computer program to generate individualized objective tests forms. *Educational and Psychological Measurement,* 39, 653-656.
Weiss, D. J., y Betz, N. E. (1973). Ability Measurement: Conventional or adaptative? *Research Report,* 73-1. Minneapolis: Psychometric Methods Program, Department of Psychology, University of Minnesota.
Weiss, D. J., y Kingsbury, G. G. (1984). Application of Computerized Adaptive Testing to Educational Problems. *Journal of Educational Measurement,* 21 (4), 361-376.
Wherry, R. J., y Gaylord, R. H. (1944). Factor patern of test items and test as a function of the corelation coeficient: Content, difficulty and constant erro factors. *Psychometrika,* 9, 237-244.
Williams, R. G., y Haladyna, T. (1982). Logical Operations for generating intended questions (LOGIC): A typology for Higher Level Test Items. En G. H. Roid y T. M. Haladyna (Eds.), *A technology for test-items writing.* New York: Academic Press.
Wulfeck, W. H., y Ellis, J. A. (1981). *Handbook for criterion-referenced testing in Navy Schools.* San Diego, California: NPRDC.
Yalow, E. S., y Popham, W. J. (1983). Content validity at the crossroads. *Educational Researcher,* 12, 10-14, 21.

RESUMEN

En este trabajo presentamos una reflexión acerca del concepto de dominio educativo como componente esencial para la construcción de pruebas de referencial criterial. Esta conceptualización, basada en principios de medición educativa, supone una revisión integrada de todos los aspectos implicados en la definición del dominio educativo, desde la definición teórica del ámbito instruccional hasta la generación de elementos, analizando sus consecuencias aplicadas. De este modo, se revisan los siguientes tópicos: la conceptualización del dominio educativo como universo de medida, sus elementos o sistemas de definición y estructuración, tecnologías de generación de *items* y sistemas de gestión de dominios definidos.

SUMMARY

In this paper, we attempt a reflexion at educational domain concept as a essential component for criterion-referenced tests construction. This conception, based on educational measurement principles, supposes an integrated review of all the facets that are implicated in educational domain definition, from theoretical definition of instructional frame to item generation, analizying its applied consequences. Thus, we have revised the following topics: the conception of the educational domain as an universe of measurement, its lements or its definition and structuration systems, the item generation technologies and management systems of defined domains.
Maritz, J. S. (1966). Smooth empirical Bayes estimation for one-parameter discrete