

El Diagnóstico de los Aprendizajes Básicos y sus Deficiencias. Un Modelo Criterial

M^a Isabel Salvador Pérez

RESUMEN:

En este artículo tratamos el problema de los modelos de diagnóstico aplicados en el campo pedagógico. Desde una perspectiva crítica, abordamos las posibilidades y limitaciones de los métodos cuantitativos o psicométricos, cualitativos y criterios, y nos definimos partidarios de la complementariedad, o de la elección de unos u otros en dependencia con la naturaleza del problema. Por esta razón, consideramos que los métodos criterios son los más idóneos para el diagnóstico de los aprendizajes instrumentales y de sus dificultades en la escuela primaria, como fundamento para adaptar a las necesidades individuales o grupales la programación de la enseñanza general o correctiva, de recuperación o de refuerzo, o las adaptaciones curriculares, condiciones básicas de la educación personalizada. Argumentamos nuestra opción describiendo el modelo de diagnóstico criterial en el que hemos trabajado y experimentado durante bastante tiempo, e incluimos datos sobre sus características, resultados y controles estadísticos que avalan su validez.

PALABRAS CLAVE: Modelos de diagnóstico pedagógico. Diagnóstico criterial. Diagnóstico de los aprendizajes básicos.

ABSTRACT:

This article discusses the issue of diagnostic models applied to the pedagogical field. We investigate the capabilities and limitations of quantitative or psychometric, qualitative and criteria methods from a critical perspective, and we favour complementarily or the choice of a method as a function of the nature of the problem. We conclude that criteria methods are the best suited for the problem of the diagnosis of instrumental learning in primary school and its problems, as a foundation for the adaptation of the planning of general or corrective, remedial or refresher education or curricular adaptations to the individual or group requirements; basic conditions of personalized education. We support our conclusion by describing the criteria diagnostic model we have developed and tested over a long period of time, and we include data on its characteristics, results and statistical controls that demonstrate its validity.

KEY WORDS: Models of pedagogical diagnosis. Basic learning diagnosis. Criteria diagnosis.

INTRODUCCIÓN

A lo largo de nuestra experiencia profesional como orientadora escolar durante casi veinte años, constatamos que muchos escolares fracasaban en la escuela sin más carencias que unas técnicas de base insuficientes, arrastrando errores y deficiencias de los aprendizajes instrumentales desde los primeros cursos de la escolaridad. Hoy, muchos años después, lamentamos la persistencia de este problema, a pesar de las disposiciones oficiales que se han venido dando para erradicarlo, dictando normas legales (clases de apoyo, de recuperación de refuerzo) que no siempre llegan a cumplirse, o no se ajustan a las dificultades particulares y concretas de cada individuo o grupo, como exige una enseñanza personalizada, o se aplican tardíamente, cuando ya son poco eficaces.

Esta cuestión nos preocupó siempre. Por este motivo, formando parte del equipo del profesor García Yagüe, iniciamos una serie de trabajos de investigación encaminados a buscar un nuevo modelo de diagnóstico y nuevos instrumentos para controlar diferencialmente los aprendizajes básicos en los primeros años de la escolaridad primaria, detectar errores y deficiencias, y corregirlos, antes de que afectaran a los aprendizajes posteriores. Los modelos convencionales, cuantitativos, válidos para describir grupos, no nos aportaban la información detallada de las situaciones singulares que necesitábamos para diseñar programas correctivos.

En este artículo comenzamos analizando las posibilidades y limitaciones de los modelos cuantitativos, cualitativos y criterios, para cumplir las funciones del diagnóstico educativo, especialmente la preventiva o de pronóstico; a continuación, describimos el modelo criterial que hemos experimentado en el diagnóstico de los aprendizajes básicos en la enseñanza primaria y sus deficiencias, y el procedimiento seguido para su elaboración; y terminamos haciendo un esbozo de dos de las investigaciones que hemos realizado con este modelo para introducir algunos controles estadísticos sobre su validez, comentando de paso ciertos penosos resultados.

1. MODELOS DE DIAGNÓSTICO. POSIBILIDADES Y LÍMITES EN EDUCACIÓN

La metodología más conveniente para llevar a cabo el diagnóstico de las situaciones educativas ha sido desde hace tiempo motivo de controversia y confrontación. El debate se ha producido más en el plano teórico o ideológico que en el práctico. La mayoría de los especialistas del diagnóstico pedagógico admiten hoy que todos los métodos pueden ser útiles y complementarios, siendo más recomendable aplicar unos u otros, o varios, según la naturaleza del problema y los objetivos que se persigan. Seguidamente vamos a recapitular los rasgos más característicos de los modelos cuantitativos, cualitativos y criterios, desde la perspectiva de su utilidad para la actividad diagnóstica educativa.

a) Diagnóstico cuantitativo o psicométrico.

La aplicación de esta metodología en el diagnóstico educativo tiene su origen en los trabajos de Galtón, Thorndike y Binet. Básicamente consiste en representar con

números los datos psicopedagógicos y tratarlos mediante cálculos estadísticos con fines científicos. Se fundamenta en el supuesto teórico de que la mejor forma de analizar la situación intelectual y educativa de un sujeto es controlar sus aciertos en una amplia gama de estímulos; cada acierto o respuesta satisfactoria se puede tomar como unidad de medida y sumarse a las otras para dar una representación final de la cualidad analizada, pero, para conocer su significación, es preciso comparar el número total de aciertos de un sujeto con otro número que le sirve de referente, recurriendo a la edad cronológica (Cocientes: intelectuales, pedagógicos, de desarrollo, etc.), o a la distribución de las puntuaciones obtenidas por la población a la que pertenece el sujeto (Baremos: percentiles, puntuaciones típicas z y sus derivadas, S ; C.I.D.; T , etc.).

Entre las muchas críticas que se han hecho, y se hacen, a estos modelos, destacamos alguna importante que hemos de tener en cuenta al realizar o interpretar un diagnóstico obtenido con esta metodología. Por ejemplo, en el caso de los *cocientes intelectuales, u otros*, a pesar de su enorme difusión, se discute y duda de que la edad cronológica sea un buen indicador para determinar un nivel de normalidad o de anormalidad, puesto que la evolución de un sujeto a lo largo de los años, en cualquiera de sus dimensiones, ni es uniforme en ritmo, ni es homogénea, ni es simultánea en todas sus manifestaciones, y, por otra parte, en los tests que recurren a esta interpretación se suele dar el mismo peso a ítems que estimulan funciones diferentes, y con diferentes grados de dificultad para diferentes sujetos. En cuanto al sistema de valoración por *baremos*, no se puede obviar que son medidas de probabilidad, y, aun suponiendo que los estudios previos estén bien hechos, sólo indican la posición del individuo dentro de un grupo normativo, por lo que deben ser utilizadas e interpretadas con mucha prudencia y mucha cautela; a lo más que se puede llegar con estos modelos es a sugerir qué sujetos necesitan una exploración individual, en profundidad, por la extremosidad de su situación en la distribución normal de su grupo de población.

García Yagüe (1987: 144-45) describe así el problema de estos métodos: “*Hacer descripciones e inferencias individuales desde datos cuantitativos es impropio e inadecuado. Las situaciones atípicas no pueden ser analizadas y descritas desde una gama de estímulos representativos de la normalidad de las variables que analiza...Clasificar a los individuos (superdotado, deficiente o inadaptado) por criterios meramente estadísticos y tomar decisiones en función de ellos sobre formas de aprender y vivir es arriesgado, peligroso y de consecuencias imprevisibles*”. Si nos apoyamos en estos modelos para definir situaciones excepcionales que pueden condicionar el futuro de personas singulares, porque son fáciles, rápidos de aplicar y cómodos de interpretar, los resultados son graves e injustos. Como dice este autor: *No se puede equiparar anormal estadístico en la distribución de una muestra normal, y anormal psicológico*. En un escrito anterior (1972: 148), ya había expresado su preocupación ante esta costumbre tan generalizada en el diagnóstico educativo y psicológico, afirmando que: “*Para que una persona sea definida cuantitativamente como genio o deficiente hay que localizar comportamientos específicos que trasciendan su condición de ser el peor o el mejor de un grupo normal*”.

Cuando en la práctica profesional de la orientación escolar hemos utilizado estos métodos nos han resultado útiles, pero insuficientes. Sólo nos dan indicadores de posi-

ción (*destacado, medio o bajo*) en una variable o en el total de una prueba que estimula tal o cual cualidad, según el número de respuestas satisfactorias dadas, pero nada informan sobre la significación de las respuestas no dadas o de los errores, o sobre las causas que originan esos fallos, no pueden definir categorías o niveles diferenciales o modalidades de aprendizaje, ni señalar puntos sintomáticos, ni definir estructuras singulares. Para aproximarnos a la personalización en la enseñanza, en la evaluación o en la orientación escolar, es decir, en la educación, necesitamos tener un conocimiento de cada individuo, de sus posibilidades y limitaciones y, sobre todo, de las causas que originan estas últimas, con el fin de tratarlas convenientemente, de forma individualizada. Pero para este tipo de diagnóstico habremos de recurrir a otros métodos.

Sin embargo, reconocemos que los métodos cuantitativos y los tratamientos estadísticos también han contribuido a mejorar algunos aspectos del diagnóstico educativo. Por ejemplo: se han delimitado estadísticamente un gran número de variables en psicología y pedagogía; se han validado los tests y sus ítems; se han orientado a millones de estudiantes al llegar a un nivel crítico, en el que se deben tomar algunas decisiones importantes, y se han establecido pronósticos aprovechables para el establecimiento de niveles de exigencias o la previsión de áreas de dificultad o de conflicto o para la elección de tareas. Asimismo, estos modelos han hecho posible el diagnóstico o la evaluación objetiva de instituciones y programas; pero su abuso, la obsesión por la medición, la “cuantofrenia”, fue en gran parte el motivo de su desprestigio.

b) Los modelos cualitativos en el diagnóstico educativo.

Estos modelos aportan análisis y descripciones muy detalladas sobre casos y situaciones singulares. Bajo su denominación se agrupan diversos procedimientos, a menudo enfocados hacia el *diagnóstico clínico*, que tiene como objetivo la interpretación de síntomas de desarrollo o el análisis de patologías, o el *diagnóstico etnográfico*, con el que se trata de comprender los fenómenos sociales desde la propia perspectiva del o de los actores.

A diferencia del enfoque psicométrico (nomotético), indagador de las regularidades psicológicas que aparecen en la generalidad de los sujetos, el método clínico (ideográfico) se interesa por la individualidad, un solo sujeto o grupo, pero estudiado en profundidad, tratando de averiguar sus particularidades y de enmarcar sus conductas en un contexto particular y concreto que les da significado. En el método etnográfico se trata de describir la realidad, descripciones detalladas de situaciones, sucesos, sujetos o interacciones, mediante el riguroso registro de las conductas observadas; realidad única, irrepetible y dinámica que “*se materializa en un continuo devenir o flujo de conducta*” (Anguera, 1985: 138)

Kelly, en 1967, publica una relación de las características comunes entre los seguidores de los modelos cualitativos, tanto en la actividad diagnóstica como en la investigación, cuyo principio rector es la libertad: en la elección de instrumentos específicos para cada sujeto; en la recopilación de los datos, ajustada a cada sujeto; en la recogida y/o registro de los datos; en el procesamiento de datos; en la asignación de pesos,

tanto en las observaciones realizadas como en la categorización de las conductas; en la provisión de pesos distintos para los mismos datos en sujetos diferentes; y en la comunicación de los resultados.

Desde la perspectiva de los defensores del enfoque psicométrico, las técnicas utilizadas (la observación de campo o participante, la entrevista o los tests proyectivos) y la libertad propuesta por los seguidores del enfoque cualitativo en todos los aspectos implican fuentes de error, poniendo de relieve una serie de peligros que acechan desde la selección de los datos hasta su registro e interpretación, especialmente riesgo de subjetivismo, es decir, carencia de objetividad y validez. No obstante, Taylor (1992:22) afirma: *“El estudio de la conducta humana demanda mucho tiempo, es intelectualmente fatigante, y su éxito depende de la capacidad del investigador...Las mediciones cuantitativas son cuantitativamente precisas; las evaluaciones cualitativas están siempre sujetas a los errores del juicio humano. No obstante, parecería que vale mucho más la pena una conjetura perspicaz sobre lo esencial, que una medición precisa de lo que probablemente carezca de importancia”*.

Para hacer frente a las críticas de subjetivismo y de falta rigor en los resultados, se proponen diversas medidas de aceptación parcial de los tratamientos cuantitativos. Por ejemplo, Reichardt y Cook (1979) proponen la utilización de dos o más métodos para *triangular* la verdad subyacente, y es mejor que sean dispares para que sus sesgos sean heterogéneos; Miles y Huberman (1984) admiten que no se puede renunciar al conteo de frecuencias de acciones o de categorías “por nosotros establecidas”; o Anguera (1985:135), defendiendo que los dos métodos, cualitativos y cuantitativos, pueden ser complementarios, especialmente en el análisis de datos.

Algunos tratadistas advierten que la mayor dificultad radica en su aplicación para evaluar el rendimiento escolar; es decir, resulta complicado decidir el criterio que permita establecer categorías, organizarlas y validarlas empíricamente, teniendo en cuenta la posibilidad de diversos niveles descriptivos.

Desde nuestra particular experiencia, los modelos cualitativos son poco aconsejables como metodología única en el diagnóstico pedagógico escolar, porque son muy costosos en tiempo e imprecisos en sus resultados; además, por su escaso formalismo y su rechazo a la estandarización, resultan impracticables cuando se trata de analizar la situación de grandes grupos de escolares en momentos de promoción o cambio, para orientarles ante la toma de decisiones importantes y comprometidas, y para desempeñar las funciones propias del diagnóstico educativo, especialmente la función preventiva o de pronóstico. Sin embargo, son muy válidos como complemento de los métodos psicométricos, en el análisis de determinados problemas particulares, de un sujeto o de algún colectivo (un grupo-clase, un centro escolar, una comunidad de padres, etc.).

c) Los modelos criterios en el diagnóstico educativo

Estos modelos representan el esfuerzo por armonizar la objetividad y la posibilidad de generalización, propia de los métodos cuantitativos, con la posibilidad de des-

cripciones que detallen las situaciones singulares y concretas, propósito de los métodos cualitativos. Para el logro de estos objetivos se necesitan nuevos instrumentos de diagnóstico. Se precisan tests que evalúen las respuestas de forma diferencial, pormenorizada, pero no de manera arbitraria, libre o subjetiva, sino de acuerdo a unos criterios previamente establecidos, tests con *referencia a criterios* diferenciadores de categorías o de *interpretación de las puntuaciones con referencia criterial*.

Pophan cita a Glaser (1963) como el iniciador de esta expresión “*evaluación basada en criterios*” para diferenciarla de la “*evaluación basada en la normalidad*” y el antecedente de los primeros tests criterioles, junto con Wells Hively (1968), que usaría la denominación “*evaluación basada en dominios*”, y Osburn (1968) “*tests de universo definido*”. A partir de la década de los años setenta del siglo veinte, se declaró un creciente interés por este modelo evaluativo.

Según Pophan (1983: 134): “*Un tests basado en criterios se emplea para determinar la posición de un individuo con respecto a un dominio de una conducta perfectamente definida*”. Entiende por “*dominio de conducta*”: “*...un conjunto de habilidades y disposiciones que los examinandos ponen en juego cuando se les pide que lo hagan en una situación de examen*”, y por “*perfectamente definidos*”: “*...los esfuerzos descriptivos de un campo de conducta*”. Aunque, considera que: “*...la evaluación basada en criterios tendrá que madurar mucho antes de que podamos valernos de ella*” (p. 21).

Si bien los modelos criterioles pueden ser un recurso idóneo para el diagnóstico al servicio de la educación, por ser el más adecuado para describir las situaciones particulares de los alumnos y alumnas, y de los grupos, respecto a objetivos educativos previamente establecidos, y poder poner en marcha programas de enseñanza y de corrección, o de desarrollo, adaptados a cada realidad concreta, lo cierto es que no hay muchas aportaciones prácticas que se ajusten a este modelo, probablemente debido a las dificultades que implican los trabajos de investigación que se han de hacer para garantizar su validez, y, en definitiva, porque requieren invertir mucho tiempo en numerosos y sucesivos controles.

Uno de los ensayos más recientes lo representan los trabajos realizados por Díaz Alcaraz y García García (2004: 23-24), en los que aportan un modelo para la evaluación de las Matemáticas, al que se refieren en los siguientes términos: “*El modelo que nosotros consideramos más adecuado es el denominado criterial, que consiste en comparar a los alumnos con los objetivos programados*”. El procedimiento a seguir consiste en “*...concretar los objetivos que deseamos que consigan los alumnos durante un periodo determinado y determinar los objetivos de evaluación, recoger la información necesaria utilizando todos los instrumentos posibles y comparar lo que ha aprendido el sujeto con lo que debería haber aprendido (objetivos propuestos)*”. Los objetivos propuestos han de ser conductuales, porque los objetivos generales de cada área no pueden ser evaluados directamente “*...están expresados en términos de capacidades que no pueden valorarse directamente, es preciso establecer criterios de evaluación para cada área que expresen conductas más fácilmente evaluables*” (p. 37).

Por estas líneas y lo expresado en otros momentos de su obra, entendemos que estos autores se refieren a la evaluación como función de control de la enseñanza-aprendizaje a partir de unas “*metas previamente fijadas*” o notas de corte preestable-

cidas. No se corresponde con nuestro concepto de lo que debe ser una evaluación o diagnóstico criterial, aunque a la evaluación inicial ellos la denominen “*evaluación diagnóstica*”.

Tampoco encontramos demasiada similitud entre nuestro pensamiento sobre el método criterial y la interpretación de Stake (2006: 42). Este autor alude a la evaluación basada en criterios para diferenciarla de la interpretativa; dice así: “*La primera consiste en determinar y representar la calidad mediante el uso de números y escalas, es decir mediante criterios*”. Hasta aquí coincidimos, pero, refiriéndose al modelo criterial, añade: “*Es la evaluación cuantitativa, más objetiva, analítica y basada en criterios*”. Evidentemente, está lejos de nuestra idea. Y precisa después: “*la mayoría de los evaluadores entienden por criterio un descriptor o atributo importante, y por estándar la cantidad de ese atributo necesaria para una cierta valoración...*”; más adelante explica: “*Un estándar es una cantidad, un nivel o una manifestación de un criterio determinado que indica la diferencia entre dos niveles de mérito distinto. Es una puntuación de corte...Se puede utilizar como estándares un nivel de rendimiento concreto...pero, generalmente, el estándar forma parte de un modo de pensar basado en criterios...*” (p. 111).

En la actualidad, encontramos con bastante frecuencia la interpretación de la evaluación por criterios como una valoración del conocimiento a partir de una línea de corte o un estándar, según el porcentaje de respuestas satisfactorias que se adopte como criterio. Nada tenemos que decir al respecto. Es una manera de buscar objetividad a la evaluación, pero desde luego no coincide con nuestro pensamiento sobre el diagnóstico criterial o, en su caso, una evaluación criterial.

Nuestro concepto sobre el método criterial de diagnóstico se corresponde con la siguiente descripción: “*el análisis detallado, pormenorizado, de la calidad de una respuesta en términos de conducta observada, que se evalúa según criterios previamente establecidos definidos por categorías, indicadores de niveles o grados de dominio o de desconocimiento de la conducta estimulada; permite obtener descripciones diferenciales y singularizadas de los individuos y de los grupos, detectar tipos de errores u omisiones, casos singulares o cuadros de síntomas, que pueden servir de fundamento a las programaciones generales y/o correctivas, de recuperación o de refuerzo, control del progreso escolar; adaptaciones curriculares, especialmente en las áreas básicas del aprendizaje escolar, sin recurrir a baremos*” (García Yagüe: 1994).

2. NUESTRO MODELO PARA EL DIAGNÓSTICO CRITERIAL DE LOS APRENDIZAJES BÁSICOS.

a) Antecedentes y finalidad

Nos remontamos a la década de los años setenta cuando iniciamos la búsqueda de un método para el diagnóstico de los aprendizajes básicos que no fuera cuantitativo, sin baremos; un método que reuniera la descripción pormenorizada de las situaciones particulares que aportan los métodos cualitativos con la objetividad y la posibilidad de realizar controles estadísticos propia de los cuantitativos.

Desde el principio (1970-1975) nos preocupó, y nos preocupa aún, el diagnóstico de los conocimientos básicos y sus dificultades en los primeros cursos de la escolaridad obligatoria. Habíamos comprobado, en nuestros años de experiencia orientadora, que una de las causas más comunes del fracaso de los escolares eran las deficiencias en las técnicas instrumentales que arrastraban y acumulaban a lo largo de la carrera escolar.

Necesitábamos disponer de un método de diagnóstico y unos instrumentos (los tests que existían en el mercado no eran adecuados para nuestros propósitos) que permitieran diferenciar, empíricamente, el mayor número de unidades de aprendizaje básico y sus dificultades, y establecer niveles de dominio, con el fin de poder llevar a cabo intervenciones pedagógicas correctivas sobre puntos concretos de dichos aprendizajes. El método criterial o de *diagnóstico cualitativo objetivado conforme a criterios previamente establecidos* (como se definió en aquel momento) era el modelo que mejor se acomodaba para conseguir estos objetivos. Le consideramos el más idóneo porque podía aportar información detallada, pormenorizada, de la situación particular de los escolares en relación con los conocimientos instrumentales y, a partir de esta información se podrían elaborar programas correctivos especialmente diseñados para combatir tales deficiencias, de los individuos o de los grupos.

Los primeros ensayos se presentaron en el V Congreso Nacional de Psicología de 1976¹. Se continuó con la construcción de tests para el análisis de procesos y la validación por niveles, sin tener en cuenta baremos, y se concretaron para la investigación que dirigió García Yagüe en 1985, por encargo del M.E.C., encaminada a seleccionar los niños bien dotados de dieciséis provincias españolas². Después, se fueron depurando las pruebas experimental y estadísticamente; a ello contribuyeron: una docena de seminarios y debates sobre las pruebas y los resultados que se iban encontrando (1982-1989), en los que participaron un número importante de especialistas en educación (profesores de diferentes niveles académicos, orientadores, psicopedagogos y psicólogos escolares); la ejecución de dos tesis doctorales (1988 y 1994)³; la realización de dos proyectos de investigación financiados por el C.I.D.E. (1989-1991)⁴ y su aplicación en varias experiencias de diagnóstico efectuadas en todos los niveles de la enseñanza primaria. Estos trabajos han conducido a la creación de cinco baterías de pruebas, algunas de las cuales están ya muy consolidadas.

1 V Congreso Nacional de Psicología (Valladolid, 1976), en la Mesa Redonda sobre *Modelos Cualitativos del aprendizaje escolar*. Presidida por García Yagüe, con la participación de su equipo: A. Lázaro; J. Crespo; B. Pacheco; A. Caballero; I. Salvador; R. López Plantada; A. Vives; B. Urcullu y A. Palomino. Publicados en la Revista de Psicología General Aplicada, nº 141/2, pp. 305-6

2 Se utilizaron en la investigación que dirigió García Yagüe en 1981 para seleccionar los niños bien dotados de dieciséis provincias españolas; primera experiencia de un programa general de diagnóstico empleando tests cualitativos objetivados y pautas de corrección criterial; véase una detallada descripción de este trabajo en J. García Yagüe (1986). *El niño bien dotado y sus problemas*. Madrid: CEPE.

3 Los datos fueron tratados también en dos tesis doctorales realizadas por M^a I. Salvador Pérez (1988) y M^a A. Camina Duránte (1994) en la Universidad Complutense de Madrid; calificadas "Cum Laude" por unanimidad.

4 Los trabajos de investigación que se hicieron para el C.I.D.E. fueron: "Un modelo para la recuperación y el refuerzo de los aprendizajes básicos" y "La situación de los alumnos en los aprendizajes instrumentales a la salida de la escuela primaria y sus implicaciones curriculares".

b) El proceso de elaboración de los instrumentos creados, su descripción y sistema de valoración

A pesar de los óptimos resultados del método criterial para el diagnóstico y tratamiento de las dificultades en los aprendizajes, en nuestro país tiene escasa tradición, posiblemente debido a lo costoso que resulta en tiempo y el esfuerzo que exige, desde la puesta en marcha del proyecto: el diseño teórico, la creación de pruebas específicas, su progresiva depuración, las etapas de experimentación, los sucesivos reajustes, los tratamientos estadísticos, los controles de validez...

El procedimiento seguido en la elaboración del modelo que presentamos ha supuesto:

- 1º. Diferenciar teórica y empíricamente el mayor número de puntos de aprendizaje básico o unidades de dificultad, transformarlos en variables e indicadores, y agruparlos mediante análisis factoriales.
- 2º. Establecer criterios para evaluar niveles de aprendizaje en cada punto o unidad de dificultad, expresarlos en términos de tareas o conductas observables que representasen grados o niveles de conocimiento, desde el dominio hasta el desconocimiento, con algunas situaciones intermedias.
- 3º. Construir instrumentos apropiados para estimular las conductas indicadoras del conocimiento y poder evaluarlas diferencialmente según los criterios establecidos.
- 4º. Contrastar los resultados reales que se encontraron con los objetivos propuestos en los programas oficiales.

Las baterías de pruebas que se han creado para estimular los aprendizajes básicos y evaluarlos por niveles y criterios preestablecidos están adaptadas a cada uno de los tres ciclos de Educación Primaria, más otras dos, una para el final de la Educación Infantil (en el momento de pasar a la primaria) y otra para el final de esta etapa (en vísperas iniciar la secundaria obligatoria). La batería del nivel superior se aplica en 190/220 minutos de forma colectiva (cuatro sesiones) y de 15/20 minutos de forma individual; la del segundo nivel ocupa 95/105 minutos la forma colectiva (dos sesiones) y 6/8 minutos de aplicación individual; y la del primer nivel exige entre 100/120 minutos la modalidad colectiva y 15/20 minutos la individual. En total acumulan más dos centenares de indicadores, evaluables diferencialmente, y medio centenar de familias de indicadores afines; diecinueve indicadores agrupan de entre dos y seis ítems, para matizar más los resultados del análisis.

Los ítems que componen cada indicador pueden ser evaluados en cinco niveles, que van desde el dominio de una noción, en términos de conducta observada, hasta el desconocimiento de la misma, con tres niveles intermedios, según la situación que se describe a continuación. Algunos ítems contenidos que pueden ser mejorados durante el ciclo se evalúan sólo en tres niveles: Dominado (no hay errores o deficiencias), Insuficiente (algunos errores o deficiencias que pueden y deben mejorarse) y Mal (nulo o gravemente insuficiente).

NIVELES DE RESPUESTAS	
DOMINADO	<i>Respuestas adecuadas y sin error</i>
BIEN	<i>Aprendizaje casi dominado. Respuestas con algún error o imperfección secundaria</i>
INSUFICIENTE	<i>Empieza a dominar. Respuestas aún con bastantes errores</i>
MAL	<i>Empieza a conocer. Predominan los errores e imperfecciones</i>
NULO	<i>Desconoce o presenta incapacidad total en esta unidad de aprendizaje. No hay respuesta o está totalmente equivocada</i>

La Ilustración I muestra un ejemplo de sólo un ítem del área matemática, normas de aplicación, corrección y valoración, y una copia de la ejecución real de un escolar de 3º de Primaria, con su correspondiente valoración y consejo (el espacio disponible no permite introducir la prueba completa).

Ilustración 1

Un ejemplo: un ítem del test criterial de Matemáticas (Nivel 2, segundo ciclo de enseñanza primaria) Parte I: Noción de cantidad y número Ítem I-4 Representación de números y cifras Normas para la aplicación de ese ítem I-4. Los textos subrayados se leen dos veces. Fila nº 4...Primera caja...En ella tenéis que escribir el número <u>trescientos siete</u>...escribidlo...Caja siguiente...escribir el número...<u>veinte mil cuatrocientos nueve</u>...Caja siguiente. Ahora tienes que <u>escribir para sumar...colocando bien los números y sumando después...cuarenta y ocho más quinientos seis</u> (en aplicación colectiva dar 20'')...Caja siguiente...<u>escribir para restar...colocando bien los números...novecientos cincuenta y uno menos ochenta y siete</u> (en aplicación colectiva dar 20''). Caja siguiente...<u>Escribir para sumar después...diecisiete...mas cuatro...mas ciento treinta y seis</u> (en aplicación colectiva dar 30'') Normas para la corrección/valoración del ítem I-4 A) Se analiza en el total de números que tenga que escribir. * Orientación /inversión de los números que ha escrito: - Bien: Orientación adecuada - Insuficiente: Invierte algún número algunas veces - Mal: Invierte algún número siempre * Giros y deformaciones que dificultan la interpretación: - Bien: No hay errores - Insuficiente: Tiende a girar o deformar algún número - Mal: Deforma siempre los números haciéndolos casi ininteligibles

B) Sólo se analiza en el ítem de la fila n° 4.

* Adecuación de los números dictados en su valor absoluto, sin contar el 0:

Bien: No ha deficiencias

Insuficiente: Cambia uno o dos de los números dictados

Mal: Cambia tres o mas de los números dictados

* Adecuación a los números dictados en su valor relativo, sin tener en cuenta el 0:

Bien: No hay deficiencias

Insuficiente: Cambia la colocación relativa de uno o dos de los números dictados

Mal: Cambia la colocación relativa de tres o mas de los números dictados

Ejecución de las respuestas de un sujeto de 3° de Educación Primaria (8 años)

Representación de la **valoración de los indicadores** de esta unidad de aprendizaje.
Noción de cantidad y número

Comparación de objetos: I

Comparación de números: B I I M

Seguridad al contar: B

Representación de números y cifras: I I B B

Manejo del cero: B B B B

Interpretación: Los indicadores valorados con Insuficiente (I) deber ser reforzados; los Mal (M) han de ser aprendidos o recuperados

c) Investigaciones, controles estadísticos y validación

Nos vamos a referir sólo a dos de las investigaciones realizadas, una al comenzar tercero de Primaria y otra al terminar la etapa, sexto curso, como soporte de la descripción de las variables que componen estas baterías de pruebas y de los controles estadísticos efectuados para validarlas; aludiremos brevemente a algunos de los resultados obtenidos.

Para poder realizar los tratamientos estadísticos de validez, agrupar responsablemente los indicadores de una misma familia y obtener una visión global y numérica de ella, a cada nivel le hemos asignado un número, con valor simbólico, representativo del grado o categoría que manifiesta la calidad de la tarea realizada.

En la **investigación realizada en 3º curso de Primaria** trabajamos con una muestra de 230-347 escolares que iniciaban el segundo ciclo en una quincena de colegios públicos de Madrid capital. Pasamos pruebas de evaluación categorizada conforme a criterios sobre las áreas de lectura (oral y silenciosa), escritura (caligráfica y ortográfica) y matemáticas (operaciones, procesos y problemas), que agrupaban más de ochenta variables estadísticamente definidas (Cuadro 1). Se trataba de una investigación tipo pretest-postest, con cuatro grupos experimentales y uno de control, e intervención de refuerzo sobre los puntos de aprendizaje frágiles, desconocidos o erróneos. La implementación de los programas de refuerzo estuvo a cargo de los maestros en sus clases, en horario lectivo; los programas de intervención habían sido acordados previamente por ellos mismos, en seminarios quincenales, con el asesoramiento de especialistas externos⁵.

Para justificar la validez de las variables y las pautas de interpretación, sometimos los datos (una vez más) a los controles estadísticos⁶. En primer lugar, tratamos de averiguar la autonomía de las variables e indicadores para poder ser tratadas independientemente; se obtuvieron correlaciones inferiores a .500, en el 90% de las variables e indicadores, *lo que avala su independencia*. En segundo lugar, para comprobar la validez estadística o de constructo, y controlar la adecuada agrupación de los indicadores en familias y en áreas, realizamos un análisis factorial (rotación oblicua o directa Oblimin). Aparecen tres grandes factores (que se vienen a corresponder

CUADRO 1: Áreas, familias de indicadores e indicadores para el segundo ciclo de primaria

LECTURA ORAL	– Velocidad lectora	– Número de palabras por minuto
	– Nivel lector	– Nivel Prelector; Silábico, Léxico y Temático
	– Deficiencia y errores	– Alteraciones; Omisiones; Adiciones; Sustituciones; Unión o separación indebida; Paradas y bloqueos
	– Modulación	– Básica: puntos, comas, acentos, interrogaciones, admiraciones – Superior: expresividad, tono y ritmo
LECTURA SILENCIOSA Y COMPRENSIVA	– Rendimiento lector	– Comprensión de palabreas
	– Seguridad lectora	– Ejecución de órdenes
		– Lagunas o textos mutilados

5 Sobre el tema de recuperación y el refuerzo de los aprendizajes básicos a partir del diagnóstico criterial se puede consultar Camina Duránte, A. (1994): Experiencias de recuperación de las dificultades lectoras en 3º de E.G.B. Universidad Complutense de Madrid. Tesis doctoral. Editado en CD (tesis@publicaciones.ucm.es)

6 Para ampliar estos datos, puede consultarse Salvador Pérez, I: El modelo criterial y sus posibilidades. *Tendencias Pedagógicas*. 1995, 2, pp. 80-2.

CUADRO 1: Áreas, familias de indicadores e indicadores para el segundo ciclo de primaria (*Continuación*)

		– <i>Búsqueda lectora en narraciones, descripciones y de información</i> – <i>Lectura de cómic</i> – <i>Lenguaje telegráfico</i> – <i>Búsqueda en textos poéticos</i> – <i>Narración motivada</i> – <i>Descripción larga</i> – <i>Narración larga</i> – <i>Rendimiento lector Inferior (I) y Superior (II) y (III)</i> – <i>Seguridad Básica (I); (II) y (III)</i>
ESCRITURA	– <i>Caligrafía</i>	– <i>Fluidez gráfica</i> – <i>Tamaño de la letra</i> – <i>Forma de la letra</i> – <i>Posición e inclinación de las letras</i> – <i>Cohesión, presión y organización especial</i> – <i>Limpieza, dominio y legibilidad</i>
	– <i>Ortografía</i>	– <i>Ortografía natural</i> – <i>M antes de P y B</i> – <i>Mayúsculas</i>
MATEMÁTICAS	– <i>Procesos operativos</i>	– <i>Orden y presentación</i> – <i>Capacidad operativa</i> – <i>Mecanismos de llevarse en sumas y restas</i>
	– <i>Agilidad operativa</i>	– <i>Sumas y restas</i> – <i>Errores</i>
	– <i>Agilidad combinatoria</i>	– <i>Problemas</i> – <i>Problemas de sumar</i> – <i>Problemas de restar</i> – <i>Problemas de sumas y restas combinadas</i> – <i>Presentación de problemas</i>

con las tres áreas básicas): a) Un *primer factor* que, desde la estructura factorial, se denominó “*Lectura inteligente*”, agrupa las variables *rendimiento lector*, *nivel lector*, *velocidad lectora* y *conceptualización verbal*; la saturación de todas ellas con el factor va desde .600 a .850; las restantes tienen saturaciones modestas o inexistentes; b) El *segundo factor* es el “*Nivel matemático*”, muestra saturaciones importantes sólo con procesos *matemáticos* (.863), *agilidad operativa* (.675) y *problemas* (.607); c) Un tercer factor corresponde al “*Nivel caligráfico*” que incluye la *escritura caligráfica*, con una importante saturación (.700), siguiendo otras de menor cuantía. Se confirma la correcta adecuación de las agrupaciones efectuadas.

Los resultados confirmaron nuestras hipótesis. Escuetamente destacamos los siguientes: a) el modelo criterial que utilizamos ha resultado muy eficaz para el diagnóstico y la recuperación de los aprendizajes básicos y sus deficiencias.; b) el tercer curso de la escolaridad primaria es un momento idóneo para realizar el diagnóstico a todos los escolares, con el fin de detectar y corregir las deficiencias en las técnicas básicas, antes de que afecten negativamente a los aprendizajes posteriores; c) la implementación de programas de refuerzo (individual o en grupo) en las áreas instrumentales, por los profesores y dentro del horario lectivo, consigue mejorar notablemente el progreso de todos los alumnos de la clase; d) existen fuertes diferencias individuales y de grupos en los conocimientos básicos al comenzar el tercer curso; e) muchos escolares promocionan sin haber superado los objetivos de la etapa anterior. También, en un trabajo anterior, lamentando el excesivo optimismo de los organismos oficiales sobre los aprendizajes de las técnicas instrumentales que asignaban para el primer ciclo de la enseñanza primaria, concluíamos: *“Pero de ahí a que los niños sólo lleguen, al terminar el ciclo, al nivel medio de lectura por palabras, o estén muy deficientes en errores y modulación lectora, mecanismos de base para el cálculo o formas de comunicación oral, indica en nuestro criterio una situación preocupante; grave en los estratos más desfavorecidos socioculturalmente”*^{7,8,9}.

En la **investigación llevada a cabo en 6º curso de primaria**, partíamos de hipótesis similares a las que nos planteamos en tercero y derivadas de sus resultados; sintetizando: a) que los niveles generales de aprendizaje al salir de la escuela primaria estaban muy por debajo de lo establecido oficialmente en las áreas instrumentales, y b) que un porcentaje alto de los alumnos que terminaban esta etapa tenían tan graves deficiencias en los aprendizajes básicos que podrían bloquear las adquisiciones posteriores o la aceptación del estudio. Se aplicaron las pruebas de nuestro modelo criterial, adaptadas para esta ocasión (Cuadro 2). La muestra, cuidadosamente seleccionada, proporcionalmente y por estratos, estuvo integrada por 516/617 escolares que terminaban el curso en 22 colegios públicos, 11 centros situados en Madrid capital (representativos de nivel sociocultural alto, zonas deterioradas y cinturón de Madrid) y 11 en la provincia (elegidos entre pequeños municipios y municipios importantes). Más de una quincena de especialistas distribuidos por áreas, debatieron y planificaron la acción, la dirección y el control de los grupos auxiliares, entrenados para la aplicación de las pruebas, el tratamiento de datos y la información a sus colegios. Tras nueve meses de trabajo, distribuidos en seis fases, llegamos a confirmar las hipótesis iniciales.

En las pruebas criteriosales que se aplicaron para analizar diferencialmente la situación de los estudiantes, los contenidos de los indicadores y los criterios de evaluación tienen validez lógica o interna. Se pueden justificar desde casi todas las teorías del

7 Salvador Pérez, I. (1988). El progreso escolar en el ciclo inicial de la E.G.B. Tesis doctoral. Madrid: Universidad Complutense.

8 Para mayor información pueden consultarse estos estudios en: García Yagüe, J.; Salvador Pérez, I. y García Cordon, T.: “Perspectivas de la investigación”; En AA.VV. (1994): *Los aprendizajes instrumentales en la Educación Primaria*. Madrid: Escuela Española, pp. 41-63

9 Los componentes responsables de cada área fueron: Camina Duránte, Fernández Baroja, Martínez Belinchón y Piqueras, en Lectura; Bárcenas y Calzón, en Ortografía; Caballero, Asensi y Garrido, en Composición; Muñiz, Barbadillo y Muñoz, en Vocabulario; Aizpún, Casals y Suárez, en Matemáticas.

aprendizaje instrumental o de sus dificultades, y aparecen constantes, de forma más o menos explícita, en los programas oficiales que se han venido proponiendo para la enseñanza primaria¹⁰ y en las acciones docentes que los desarrollan.

Para la comprobación de la validez externa se han realizado repetidos análisis factoriales en diferentes cursos y en diferentes etapas, pero, como no es posible introducir en un artículo la totalidad de los estudios llevados a cabo con estas baterías, sólo hemos seleccionado algunos resultados de los tratamientos estadísticos efectuados con los datos de la investigación a la que venimos haciendo referencia.

CUADRO 2: Áreas, familias e indicadores para el final del tercer ciclo

LECTURA ORAL:	Nº DE INDICADORES	CRITERIOS DE EVALUACIÓN
– Nivel general de lectura	Cuatro	3-5 niveles
– Velocidad lectora	Uno	(p/m)
– Modulación lectora básica	Cuatro	3-5 niveles
– Modulación lectora compleja	Cuatro	3-5 niveles
– Errores lectores elementales	Cuatro	3-5 niveles
– Errores lectores complejos	Cuatro	3-5 niveles
LECTURA SILENCIOSA:		
– Búsqueda en texto sencillo	Uno	3-5 niveles
– Texto e ilustración	Uno	3-5 niveles
– Búsqueda en texto condensados	Uno	3-5 niveles
– Búsqueda en texto literario	Uno	3-5 niveles
– Resumen/interpretación en texto largo	Uno	3-5 niveles
– Precisión lectora		(estadístico)
ESCRITURA:		
– Fluidez gráfica	Dos (2)	3 niveles
– Precisión en representación de letras	Cuatro	3 niveles
– Precisión en unión letras/palabras	Cuatro	3 niveles
– Diferenciación de conjuntos	Cuatro	3 niveles
– Tamaño/espacio letra/palabra	Cuatro	3 niveles
– Representación posicional	Cuatro	3 niveles
– Dominio gráfico	Cuatro	3 niveles
– Forma de presentación	Cuatro	3 niveles
ORTOGRAFÍA:		
– Natural	Ocho	3 niveles
– De reglas básicas	Cuarenta y uno	3 niveles
– Visual/mecánico	Diez	3 niveles

10 Un interesante estudio sobre esta cuestión puede encontrarse en A. Camina (1996). *La lectura en los programas de la enseñanza básica española*. Madrid: Departamento de Didáctica y Teoría de la Educación. Universidad Autónoma de Madrid.

CUADRO 2: Áreas, familias e indicadores para el final del tercer ciclo
(Continuación)

– Acentuación visual	Seis	3 niveles
– Acentuación compleja	Cinco	3 niveles
– Puntuación usual	Siete	3 niveles
– Puntuación compleja	Dos	3 niveles
MATEMÁTICAS:		
– Escritura y reconocimiento de números	Uno	4 niveles
– Ordenación de decimales y fracciones	Uno	4 niveles
– Operaciones enteros decimales	Uno	6 niveles
– Operaciones con fracciones	Uno	7 niveles
– Equivalencia de fracciones	Uno	2 niveles
– Búsqueda común denominador	Uno	2 niveles
– Problemas del segundo ciclo	Uno	4 niveles
– Problemas del tercer ciclo	Uno	2 niveles
– Agilidad manipulativa básica	Uno	2 niveles
– Agilidad manipulativa con decimales y fracciones	Uno	3 niveles

VOCABULARIO: FAMILIAS DE INDICADORES CRITERIOS de EVALUACIÓN

– Sinónimos	Uno	(estadísticos)
– Antónimos	Uno	(estadísticos)
– Palabras compuestas	Uno	(estadísticos)
– Palabras polisémicas	Uno	(estadísticos)
– Asociación de significados	Uno	(estadísticos)
– Elementos formales	Uno	(estadísticos)
– Verbos irregulares	Uno	(estadísticos)
– Categorías gramaticales	Uno	(estadísticos)
– Familia de palabras	Uno	(estadísticos)
– Precisión asociativa	Uno	(estadísticos)

VOCABULARIO: FAMILIAS DE INDICADORES CRITERIOS de EVALUACIÓN

COMPOSICIÓN:

– Presentación y caligrafía	Cuatro	3 niveles
– Ortografía	Cuatro	3 niveles
– Concordancia y sintaxis	Cuatro	3 niveles
– Vocabulario	Cuatro	3 niveles
– Organización del escrito	Cuatro	3 niveles
– Estilo/forma de expresión	Cuatro	3 niveles
– Creatividad	Cuatro	3 niveles

El control estadístico del grado de autonomía de las familias de indicadores de las pruebas de este nivel superior, de las agrupaciones que pueden organizarse con ellas y la conveniencia o no de adscribir las a las grandes áreas de aprendizaje tradicional se llevó a cabo con 44 familias de indicadores, habiendo realizado 14 matrices de correlación y 12 análisis factoriales¹¹. Llegamos a las siguientes conclusiones:

1ª. *Las variables y las familias de indicadores demuestran que existe entre ellas una alta autonomía y pueden ser tratadas independientemente.* Un tercio de las correlaciones (32%) dan índices inferiores a .200; la mayoría (82%) queda por debajo de .400; sólo muestran una relación alta, importante, la lectura oral con la modulación lectora (.754), la ortografía mecánica con la ortografía de acentos (.746) y el estilo de composición con vocabulario (.751).

2ª. *Las familias de indicadores de la misma área tienen relaciones estadísticas significativas entre ellas, justificándose su agrupación.* Casi todas las familias de indicadores de lectura oral, caligrafía, composición, ortografía, vocabulario y matemáticas aportan correlaciones iguales superiores a .400 con las de su área. Los análisis factoriales también confirmaron estos resultados; la matriz rotada (rotación Oblimin) de 33 variables ratificó los agrupamientos por áreas de las familias de indicadores de matemáticas, ortografía, caligrafía (sin fluidez), composición (sin presentación) y lectura oral.

Como podemos apreciar los resultados de los controles estadísticos son muy parecidos a los obtenidos en la investigación realizada en el segundo ciclo. Concluimos, por lo tanto, *que se puede confiar en la validez lógica y estadística de este modelo de diagnóstico y de las baterías de pruebas que utiliza.* Aunque, en este nivel superior, hay aún algunas áreas que pueden ser mejoradas (matemáticas, vocabulario y composición) para acomodarse al modelo criterial que proponemos.

En cuanto a los resultados del estudio descriptivo de la situación de los escolares al terminar el 6º curso, comprobamos que los niveles del aprendizaje básico, después de seis años de permanencia en la escuela, era muy deficitario. Es frecuente leer en las conclusiones de los directores de grupo-área expresiones como estas: “escaso dominio”, “importantes deficiencias”, “rendimiento medio insuficiente”, “resultados peores de los deseados”, etc.; y se lamentan de que en este curso, relativo a estos aprendizajes instrumentales, “se presta poca atención”, “hay despreocupación”, “desconocimiento de estrategias didácticas eficaces [área ortográfica]”, “es una actividad para rellenar espacios”, “falta de un planteamiento didáctico o metodológico [en composición escrita]”, “están iniciando el aprendizaje de la lectura aplicada al estudio; tienen un nivel muy deficitario”, etc.

Si esto sucede en el momento crítico de pasar a la enseñanza secundaria obligatoria, respecto a conocimientos instrumentales necesarios para adquirir otros aprendizajes posteriores, se podía prever fácilmente el gran fracaso de muchos estudiantes en la

¹¹ Pueden consultarse los resultados completos de estos tratamientos estadísticos en García Yagüe, J.; Salvador Pérez, I y García Cordon, T. en la obra citada, pp. 51-54.

nueva etapa, y su dramática repercusión en dimensiones tan importantes como la autoestima, la trayectoria vital y/o los problemas de conducta individual o social. Lamentablemente, el tiempo ha confirmado la catástrofe que ha supuesto la E.S.O. Y el problema sigue sin resolver.

CONCLUSIONES:

Hemos comprobado que el diagnóstico criterial de los aprendizajes básicos aplicado a todos los escolares en los momentos de cambio (cambio de ciclo o cambio de etapa) hace posible adaptar la enseñanza a las situaciones particulares de los individuos y de los grupos, pronosticar dificultades y prever fracasos, que se podrán evitar a tiempo. El diagnóstico realizado exclusivamente a los individuos cuando ya sufren problemas o dificultades empobrece la educación de todos.

El modelo de diagnóstico criterial que hemos descrito tiene en cuenta las respuestas erróneas, las lagunas u omisiones o las respuestas originales, por lo que permite detectar cuadros de síntomas de dificultades y procesos singulares. Nos ha resultado idóneo para conseguir descripciones pormenorizadas de la situación de los escolares, individuales o de grupos, diferenciando niveles de dominio de unidades básicas y puntos deficitarios en el aprendizaje de las áreas instrumentales, con lo que se facilita la organización de programas específicos para su debido tratamiento correctivo, antes de que se conviertan en obstáculos en el aprovechamiento escolar. Las experiencias de recuperación que se han llevado a cabo en este sentido han dado resultados muy satisfactorios.

Finalmente, consideramos que *en el campo pedagógico, los modelos de diagnóstico criterial, con validez demostrada mediante controles estadísticos, son los más útiles para el análisis diferencial de los aprendizajes, como fundamento para la programación de la enseñanza general, correctiva, de recuperación o refuerzo, y para argumentar de forma responsable la toma de decisiones en situaciones educativas difíciles; exigencias básicas cuando se pretende mejorar la calidad de la enseñanza y llevar a la práctica la educación personalizada.*

BIBLIOGRAFÍA CITADA

- Anguera, M^a T. (1985). *Metodología de la observación en las Ciencias Humanas*. Madrid: Cátedra.
- AAVV (1993). El Nivel de los Aprendizajes Instrumentales al Terminar la Escuela Primaria Española (11/12 años) y sus Implicaciones Pedagógicas. *Tendencias Pedagógicas* (0), Monográfico dedicado a las conclusiones del trabajo dirigido por el Dr. García Yagüe con la subvención del CIDE. Edita el Departamento de Didáctica y Teoría de la Educación (U.A.M.).
- AAVV (1994). *Los aprendizajes instrumentales en la Educación Primaria*. Madrid. Escuela Española.

- Camina Duránte, A. (1996). *La lectura en los programas de la enseñanza básica española*. Madrid: Departamento de Didáctica y Teoría de la Educación (U.A.M.).
- Díaz Alcaraz, F. y García García, J. J. (2004). *Evaluación criterial del área de Matemáticas*. Barcelona: CISSPRAXIS.
- García Yagüe, J. y otros (1987). *Diagnóstico Pedagógico y Técnicas de la Orientación*. Madrid: UNED.
- García Yagüe, J. (1986). *El niño bien dotado y sus problemas*. Madrid: CEPE.
- Glaser, R. (1963). Instructional Technology and the Measurement of Learning Outcomes: Some Questions. *American Psychologist* (18), 519-21.
- Hively, W. (1974). Introduction to Domain-Referenced Testing. *Educational Technology* (14), 5-10.
- Kelly, E. L. (1967). *Assesment of human characteristics*. Wadsworth Publising Company, Inc.
- Huberman, A. M. y Miles, M.B. (1983). "Drawing valid meaning from qualitative data: Some techniques of data reduction and display". *Quality & Quantity*, 17, 281-339.
- Osburn, H. G. (1968). "Item Sampling for Achievement Testing". *Educational and Psychological Measurement*, 28, 95-14.
- Pophan, W. J. (1983). *Evaluación basada en criterios*. Madrid: Magisterio Español.
- Reichardt, Ch. y Cook, T. D. (1982). "Beyond qualitative versus quantitative methods". En T. D. Cook y Ch. S. Reichardt (Eds.), *Métodos cualitativos y cuantitativos en investigación evaluativa*. Madrid: Morata.
- Stake, R. E. (2006): *Evaluación comprensiva y evaluación basada en estándares*. Barcelona GRAÓ.
- Taylor, S. J. y Bogdan, R. (1992). *Introducción a los métodos cualitativos de investigación*. Barcelona: Paidós.

