

# Validez de la evaluación del potencial de aprendizaje

M. Dolores Calero García  
Universidad de Granada

La evaluación del potencial de aprendizaje surge como alternativa a la evaluación tradicional de la inteligencia asociada a la evaluación de niños y adolescentes con déficits o problemas de aprendizaje. Sin embargo, en la actualidad, está recibiendo gran atención por parte de diferentes grupos de investigación internacionales, no sólo en relación a sus supuestos y su metodología, sino también a su aplicabilidad en otras poblaciones objetivo. En este trabajo se revisan tres cuestiones importantes relacionadas con su validez, que han sido muy debatidas en una serie de trabajos publicados recientemente. Éstas son: su supuesto básico de evaluación de la capacidad de aprendizaje, las diferentes alternativas para llevar a cabo la medida del cambio que implica su aplicación y los estudios de validez desarrollados.

*The validity of the learning potencial assessment.* The assessment of Learning Potential, is an assessment methodology that was born firstly, as an alternative to traditional intelligence assessment approach, asociated to children and young people with deficits or learning disabilities. Nowadays, however, Learning Potential Assessment is receiving an important attention on differents international research groups, not only in order to its aims and methodology, but also to its applicability on other populations. In this work, three important cuestiones related to Learning Potential Assessment validity, that have been discussed in several works published recently, have been reviewed. These cuestiones are, its way of performing the assessment of learning ability, the different alternatives to do the change measurement that involves its application and the validity studies that have been developed.

La evaluación del potencial de aprendizaje o evaluación dinámica, según las dos denominaciones más importantes que este tópicico recibe actualmente, constituye una metodología de evaluación que surgió hace unos 30 años como una alternativa a la evaluación tradicional de la inteligencia orientada inicialmente a la evaluación de la capacidad de aprender en niños que presentaban déficits intelectuales o de aprendizaje en los tests tradicionales de inteligencia (Fernández y Fernández, 1990).

La idea subyacente, según Resing (2001), retoma la posición de Binet y Thorndike sobre la Inteligencia entendida como capacidad de aprender, influida ahora por la teoría de Vygotski en el sentido de que no sólo se interesan por determinar el nivel intelectual, sino también por la posibilidad de instrucción del sujeto evaluado (Resing, 2001). Así, desde sus inicios, la mayoría de los grupos de investigación que han desarrollado esta tecnología (Feuerstein y colaboradores, Budoff, Haywood, Lidz, el grupo de California-Berkeley: Brown, Campione, Ferrara..., los alemanes: Wield, Tzuriel, Guthke, y en España: Fernández-Ballesteros, Calero, etc.) han asumido como constructos teóricos fundamentales el concepto de *Zona Desarrollo Próximo* de Vygotski (Vygotski, 1978) y los de *Modificabilidad Cognitiva* y *Mediación* desarrollados por Feuerstein (Feuerstein, Rand y Hoffman, 1979).

La metodología general de evaluación del potencial de aprendizaje es similar al procedimiento planteado por Vygotski para evaluar la «Zona de Desarrollo Próximo» pero modificado por los diversos autores que trabajan en este campo (Tzuriel, 2001). En general, se trata de incluir una fase de entrenamiento en la situación de evaluación. Esta fase puede ser suministrada entre dos aplicaciones de la misma prueba (según un formato de test-entrenamiento –test, o puede administrarse ítem a ítem, como respuesta a los errores cometidos por el sujeto en cada ítem aplicado. En cualquier caso se trata de una fase de mediación activa, monitorización y guía, dirigida a dotar al evaluado de aquellas habilidades básicas que, según el evaluador, le van a llevar a una mejor ejecución en ese dominio específico (Haywood, 1993; Gerber, 2001).

En estos 30 años, gran cantidad de grupos de investigación han trabajado en esta área y desarrollado diferentes instrumentos y diferentes modos de aplicación e interpretación y entre todos se han abarcado diversas poblaciones objetivo. Esto ha permitido establecer algún tipo de clasificación de técnicas de evaluación del P.A. Una de las clasificaciones más interesantes ha sido la realizada por Brown, Campione, Webber y McGilly (1992), que las divide en técnicas centradas en el producto y técnicas centradas en el proceso, pues en ella se tienen en cuenta criterios tales como: el centro de interés, el tipo de interacción y el objetivo que la evaluación pretende.

En todos los casos se han enfatizado: 1) las ventajas que esta aproximación tiene para detectar a individuos que, aunque inicialmente aparezcan como deficitarios, son susceptibles de mejora; 2) su utilidad para poblaciones de otras etnias o con déficits de diferentes tipos; 3) su capacidad predictiva superior a la evaluación

tradicional en relación a las habilidades que pueden ser entrenadas; 4) la sensibilidad a factores no intelectuales que afectan al rendimiento, tales como el autoconcepto o la motivación; y 5) el hecho de que supongan medidas de habilidad más homogéneas que la evaluación tradicional (Calero, 1995; Lozano Fernández y col., 2000; Lauchlan y Elliot, 2001). También hemos asistido a la investigación sobre su uso potencial en los contextos más diversos: poblaciones de inmigrantes (Hessels y Hamers, 1993), superdotados (Lidz y Macrine, 2001), adultos con esquizofrenia (Wiedl, Wienöbst y Schöttke, 2000), personas con daño cerebral (Schöttke, Baltram y Wiedl, 1993) y ancianos sanos y/o con demencias (Baltes, Dittman-Kohli y Kliegl, 1986; Baltes, Kühl y Sowarka, 1992; Fernández-Ballesteros y Calero, 1995, etc.)

No obstante, a pesar del optimismo con que esta aproximación evaluativa se ha desarrollado, el buen acogimiento dispensado por los psicólogos aplicados y por los educadores y el impacto positivo, a nivel social, que ésta ha tenido; su uso en contextos aplicados no ha tenido la difusión que podría esperarse. Algunos autores piensan que el tiempo que conlleva su aplicación, los conocimientos que, de los evaluadores, requiere y el hecho de que sea un área poco familiar para los psicólogos en ejercicio, son los causantes de este fenómeno (Guthke y Stein, 1996; Resing, 2001).

Sin embargo, otros autores se han cuestionado si realmente existe alguna diferencia entre la evaluación dinámica y la tradicional. En concreto Sternberg y Grigorenko (2001) plantean que el supuesto fundamental de esta área consiste en que una persona aprende en una situación de evaluación dinámica, pero no en una situación de evaluación estática o tradicional, y ésta es una diferencia «quimérica» (según sus palabras). Otros investigadores, más interesados en la medida del cambio que en la promoción del cambio, han planteado las dificultades metodológicas que supone trabajar con puntuación de ganancia (Elliot, 2001) y otro grupo señala la ausencia de estudios de validez consecuentes con la propia estructura y objetivo de estas técnicas (Lauchlan y Elliot, 2001).

Aunque creemos que muchas de estas críticas no están justificadas y que esta situación no es sino un reflejo de la inmadurez de un área de investigación que provoca todavía la gran falta de claridad conceptual que se percibe al observar, por ejemplo, cómo aún no se diferencia claramente entre algunos de los conceptos utilizados: test/evaluación, mediación/instrucción y/o aprendizaje/modificabilidad (Elliot, 2001; Kaniel, 2001); y aunque en otros trabajos ya hemos intentado responder a algunas de estas cuestiones (Fernández-Ballesteros y Calero, 1993/95/2000), vamos a intentar aportar algunos datos con el único fin de intentar clarificar los que nosotros consideramos es el estado de la cuestión.

¿Supone la evaluación dinámica una medida de potencial de aprendizaje diferente de la medida de la inteligencia que se deriva de la evaluación tradicional?

Como antes señalábamos, Sternberg y Grigorenko planteaban que cualquier situación test es una situación de aprendizaje y que por tanto ésta no es una diferencia fundamental entre la evaluación dinámica y la tradicional. Lo único que, según ellos, añade el test dinámico sobre el estático es que incorpora el aprendizaje explícito al implícito que se da en cualquier situación test, y ésta sería la diferencia apreciada entre las puntuaciones pre- y post-. A esta afirmación en su momento se le hicieron varias puntualizaciones que creemos interesante retomar.

En primer lugar, no debe confundirse la evaluación dinámica con un test dinámico, pues en la aproximación dinámica no se busca aumentar la ejecución del sujeto en el posttest (aunque éste sea el resultado usualmente visible), sino que lo que se busca es controlar lo que comprende el examinado a fin de suministrarle ayudas que le lleven a tener las mismas oportunidades que tienen otros de adquirir un repertorio (Haywood, 2001). O sea, es la situación la variable predictora de las futuras ejecuciones (Tzuriel, 2001); como hemos dicho en otro momento se trata de hacer micro-experimentos en los que «*la mejora que se consigue ha de ser considerada como una muestra del comportamiento en situaciones de aprendizaje semejantes*» (Fernández-Ballesteros, Calero, Campillonch y Belchí, 1987/90; p. 2).

Tampoco creemos que deba oponerse, como ellos hacen: estático a dinámico. La característica fundamental de la evaluación tradicional frente a la del P.A. no es que sea estática, es que es estándar: igual para todos los sujetos evaluados, uniforme y orientada hacia criterios normativos, mientras que «dinámico» se refiere a «adaptado al sujeto» (Kaniel, 2001).

Sternberg y Grigorenko, además, pierden de vista el objetivo fundamental de la evaluación del P.A., pues es claro que todos aprendemos, pero no todos aprovechamos en el mismo grado, ni a la misma velocidad, las señales que se nos suministran; esto es, la diferencia de esta evaluación con la tradicional está en que detecta la capacidad de aprovecharse del entrenamiento en aquellos sujetos que desde una evaluación estándar se supondrían con déficits y/o problemas de aprendizaje (Tzuriel, 2001).

Una tercera cuestión, quizá la más importante, es que existe evidencia empírica de las mejoras significativas de ejecución obtenidas por los sujetos evaluados mediante procedimientos dinámicos.

Gran cantidad de autores han mostrado datos que reflejan diferencias significativas entre sujetos, que son entrenados como parte del procedimiento en una técnica en particular, frente a grupos en los que simplemente se da una evaluación repetida. Esto se manifiesta con diferentes técnicas y sujetos, por ejemplo, Lidz y Thomas (1987) con el PLAD muestran ganancias significativas en grupos experimentales y controles en niños de 3 a 5 años. Tzuriel y Klein (1987) muestran que la pura repetición no mejora la ejecución y el entrenamiento sí. Baltes, Dittman-Kohli y Kliegl (1986) aportan datos demostrativos de mejoras significativas en muestras de ancianos utilizando tests como el *test de factor G* de Cattell y el subtest espacial del PMA y Carlson y Wiedl (1979) informan sobre mejoras significativas en escolares con el Raven Matrices Progresivas.

Para mostrar mejor este efecto en la tabla 1 recogemos algunos datos obtenidos por nosotros en diferentes investigaciones, con diferentes grupos de sujetos y técnicas.

Por ejemplo, el test de matrices progresivas de Raven –que tiene una fiabilidad test-retest de .91, que implica una variabilidad menor de 3 puntos directos en medidas repetidas (Raven, 1969)–, aplicado según un formato test-entrenamiento-test, ha alcanzado unas ganancias medias de 11.11 con un rango entre 0 y 20 puntos; siendo éstas similares en los diferentes grupos muestrales con los que hemos trabajado: niños con déficit intelectual, adolescentes y ancianos.

En el PWG: *Picture Word Game* (Calero y Márquez, 1998) la puntuación inicial de cualquier grupo de niños sería 0, pues se trata de una tarea desconocida, sin embargo tras el entrenamiento niños preescolares y niños con síndrome Down alcanzan puntuaciones similares que se sitúan en torno a los 20 puntos (Robles y Calero, 2003).

<i>Tabla 1</i>					
Mejoras obtenidas en diferentes investigaciones sobre técnicas de potencial de aprendizaje (* se indica media de ganancias simple (post-pre) y entre paréntesis desviación típica)					
Investigación	Año	N	Edad	Características	Ganancias medias *
<b>Sobre EPA (M.P. Raven)</b>					
Fernández-Ballesteros y Calero	1993	48	10-14	Deficiencia mental	5.19 (5.20)
Calero	1994	156	13-16	Dificultades de aprendizaje	9.93 (6.02)
Calero y Lozano	1994	18	60-80	Ancianos	11.11 (6.64)
Fernández-Ballesteros y Calero	1995	25	58-83	Ancianos sanos	9.54 (6.04)
<b>Sobre PWG: Picture Word Game</b> (codificación simbólica)					
Calero y Márquez	1998	227	5-6	Preescolares	19.9 (8.2)
Robles y Calero	2003	33	2-22	Síndrome Down	20.18 (6.04)
<b>Sobre AVLT: Auditory Verbal Learning Tests</b> (Memoria Verbal)					
Calero y Navarro	2002	123	60-85	Ancianos sanos	3.97 (2.00)

En el AVLT: *Auditory Verbal Learning Tests*, las ganancias medias máximas obtenidas por ancianos de 60 y 70 años, según barremos (Strauss y Spreen, en Lezak, 1995), se sitúan en 2.5 puntos, mientras que en la versión de potencial de aprendizaje los resultados obtenidos muestran una media de ganancia de 3.97 (Calero, Navarro, Arnedo, G<sup>a</sup>-Berben y Robles, 2000).

Evidentemente, para determinar con garantías si en las técnicas de P.A. se da o no un aprendizaje significativo debemos de establecer, en primer lugar, el nivel de confianza que debe aplicarse al cambio de ejecución sufrido por un individuo en la situación de evaluación y ello nos lleva directamente a la siguiente cuestión. No obstante, tengamos en cuenta uno u otro criterio, está claro que estas técnicas sí detectan la capacidad de aprender de diferentes tipos de sujetos que en la evaluación tradicional aparecen como deficitarios.

¿Cómo debe medirse el cambio producido en un individuo por aplicación de una técnica de potencial de aprendizaje?

Sin entrar en las propuestas hechas desde la teoría del Rasgo Latente, dado que aún no están suficientemente desarrolladas para su utilización en este contexto (May y Nicewander, 1998), existe un considerable número de trabajos que analizan qué tipo de puntuaciones son las más convenientes para medir los efectos de tratamiento. Gran cantidad de autores trabajan con puntuación de ganancia simple, y algunos la prefieren a la puntuación posttest cuando la diferencia de media entre los grupos pre y post es significativa (May y Nicewander, 1998). No obstante, comparando distintos tipos de puntuaciones, algunos autores (Cronbach y Furby, 1970) plantearon que la puntuación de ganancia simple (post-pre) no era fiable por dos razones principales: los problemas de fiabilidad y validez de las puntuaciones de ganancia, pues éstos dependen de la fiabilidad y validez del pretest y posttest, y los problemas de escalamiento y efecto de la regresión derivados del hecho de que el pretest del sujeto forme parte de su cálculo. La puntuación de ganancia residual mejora algo esta situación, pero no demasiado, ya que sigue dependiendo de las puntuaciones pre y post (Embretson, 1987).

Así, para algunos autores, trabajar con el posttest no presenta problemas si las normas del test son adecuadas (Embretson, 1987)

y de hecho esta opción se utiliza bastante y en un considerable número de trabajos se señala que la puntuación posttest es un buen indicador de potencial de aprendizaje (Lauchlan y Elliot, 2001; Resing, 2001) que aumenta el valor predictivo de estas técnicas muy por encima de las ganancias residuales. En concreto, Resing (1997) encuentra una fiabilidad del posttest de entre .90 a .95 frente a un .63 a .85 de las ganancias residuales.

Otra opción ha sido el algoritmo de Baltram, Schöttke y Wiedl (1993). Éste se basa en la fiabilidad del pretest y consiste en estimar mediante análisis de regresión lineal la ejecución hipotética de un sujeto en un test paralelo. Se trata de determinar en primer lugar si ha ocurrido un cambio significativo en un sujeto y para ello establecen un intervalo de confianza con el que comparan la puntuación obtenida por él en el posttest. Así, aunque se utilice una diferencia pre-post, esta regla de decisión limita la fuentes de error debidas a los efectos suelo y techo del test.

Siguiendo un razonamiento similar, nosotros hemos utilizado un cálculo sobre la media y desviación típica del pretest para considerar cuándo una ganancia es significativa. Este criterio arroja resultados similares al algoritmo anterior y es de más fácil aplicación. Según éste, se considera que una persona es alto-puntuadora cuando su puntuación pretest está en un rango igual al techo de la prueba menos 1.5 desviaciones típicas del pretest de su grupo. Asimismo, se considera que una mejora es significativa cuando la ganancia simple es mayor de 1 desviación típica del pretest. Esto supone categorizar a los sujetos evaluados (establecer categorías en una dimensión continua utilizando un punto de corte), pero implica fijar un criterio de valoración de las puntuaciones de los tests de aprendizaje que limita los errores metodológicos derivados del trabajo con puntuaciones de ganancia y elimina los errores debidos a ambos efectos: techo y suelo del test (Dwyler, 1996). Implícitamente significa aceptar una metodología de clasificación, cuestionada por algunos investigadores, pero que indudablemente ha tenido importantes implicaciones en el uso pedagógico y diagnóstico de esta aproximación (Lauchlan y Elliot, 2001). Una vez que se establece que una persona tiene una ganancia significativa se puede trabajar con la puntuación de ganancia considerándola como continua.

### Evidencias de validez de las técnicas de evaluación del potencial de aprendizaje

Como señalan Carlson y Wiedl (2000), para demostrar la validez en este contexto se requieren diferentes tipos de evidencias.

En primer lugar habría que demostrar la validez de constructo y en este sentido, a pesar de las críticas actuales, tradicionalmente se ha supuesto que la evaluación del potencial de aprendizaje poseía validez de constructo, puesto que con ella se demostraba la modificabilidad de la habilidad de partida en un sujeto tanto en el test entrenado como en tests que evaluaban la misma habilidad.

Embretson y Prenovost (2000) realizan un trabajo en el que intentan comprobar la validez de constructo entendida en dos aspectos separados (la representación del constructo y el espacio nomotético) de dos instrumentos de potencial de aprendizaje –uno espacial y otro de razonamiento–, frente a tests estáticos y considerando variables afectivas y motivacionales que pueden influir en la ejecución. Encuentran, mediante análisis factorial confirmatorio, que existen diferencias entre dominios y entre medidas estáticas y dinámicas y que la evaluación dinámica es más sensible a aspectos no intelectivos.

En una serie de trabajos realizados hemos intentado mostrar la validez de constructo y predictiva de estas técnicas al comprobar cómo los sujetos (niños con deficiencia mental) que mejoraban significativamente en la *EPA: Evaluación del Potencial de Aprendizaje*, eran los que mejoraban significativamente en un programa de entrenamiento cognitivo, encontrando una sensibilidad de .78 y una especificidad de .79 (Fernández-Ballesteros y Calero, 2000). En otros trabajos, esta vez con ancianos, hemos mostrado que utilizando dos diferentes técnicas de potencial de aprendizaje aquellos ancianos que aparecían como ganadores en una de ellas mejoraba también en otra, independientemente de otros factores tales como la habilidad cognitiva de partida. La tabla 2 refleja la distribución encontrada entre dos tests de potencial de aprendizaje para un grupo de ancianos (N= 60), en este caso las pruebas utilizadas eran el test de Wisconsin en su versión potencial de aprendizaje y el AVLT- P.A., o sea, eran pruebas que evaluaban diferentes habilidades y que, sin embargo, utilizaban un procedimiento de aplicación similar de test-entrenamiento-test, mostrando en su análisis la posibilidad de mejora en ambas pruebas (Calero y colaboradores, 2000). Parece, pues, que la validez de constructo está suficientemente demostrada.

Otros autores han intentado comprobar la validez de los tests de potencial de aprendizaje en relación con la predicción de éxito académico (para una revisión Carlson y Wiedl, 2000) –trabajos cuestionados por algunos autores (Lauchlan y Elliot, 2001; Stern-

berg y Grigorenko, 2002), puesto que utilizaban una medida dinámica con un criterio estático de cuál querían alejarse–; también han sido relacionados con éxito con el rendimiento obtenido en programas de entrenamiento específicos, tanto de problemas formales (Budoff, Meskin y Harrison, 1971), como de solución de problemas cotidianos (Wiedl, Wienöbst y Schöttke, 2000), encontrando, en ambos casos, buena capacidad predictiva de las técnicas de P.A. con respecto al aprendizaje obtenido.

En España, Orrantía, Morán y Gracia (1998) han comprobado la validez incremental de las medidas dinámicas sobre las estáticas, encontrando que mientras las medidas estáticas sólo explican el 0,31 de la varianza de la ejecución, sumar a la evaluación una medida dinámica aumentaba el nivel explicativo al 0,65.

### En resumen

Si bien queda mucho por investigar en esta aproximación evaluativa aún tan joven, y todavía existen matizaciones que pueden hacerse sobre supuestos, sobre métodos y sobre posibilidades de investigación, creemos que, a luz de los resultados obtenidos, la evaluación del potencial de aprendizaje ocupa un lugar propio y diferenciado en la evaluación de la inteligencia.

Muchos psicólogos y educadores han señalado la inadecuación de los tests tradicionales para revelar la capacidad de aprender en niños con dificultades de aprendizaje o en aquellos en los que factores no intelectuales son la causa de sus fallos (Tzurriel, 2001). Los tests tradicionales se eligieron según parámetros estáticos que impidan el aprendizaje en esa situación. La evaluación del potencial de aprendizaje, a pesar de sus limitaciones, tiene, pues, un objetivo clínico que la evaluación tradicional no cubre.

Su objetivo no es medir tan sólo la ejecución de los sujetos, sino su posibilidad de aprendizaje.

Su fin no es el pronóstico académico, entendido del modo tradicional, sino la estimación de la posibilidad de aprovechamiento de diferentes programas de entrenamiento cognitivo.

Su población diana fundamental es aquella susceptible de mejora, o sea, que esté en desarrollo o en declive: porque sea joven, porque sea deficitaria, porque sea mayor, etc.

Su campo de actuación no es sólo la Psicología educativa, sino también la clínica: prescripción y pronóstico; la Psicología del trabajo: predicción, orientación y selección profesional; y la Psicología social: transcultural.

Así, hoy se investiga, y se puede llevar a cabo, su uso:

- Como alternativa a la evaluación tradicional de la inteligencia en aquellas personas con problemas de comunicación debidas a diferencias culturales, o a déficits sensoriales o a privación sociocultural.
- Con fines predictivos del aprendizaje cuando lo que se intenta es estimar el aprovechamiento que un sujeto o grupo tendrá de un programa de entrenamiento específico, sea éste educativo, rehabilitador (en estas ocasiones a veces se denomina evaluación del potencial de rehabilitación) o profesional.
- Con fines prescriptivos o pronósticos en aquellos sujetos o grupos en los que lo que interesa es juzgar si podremos llevar a cabo entrenamientos para aumentar o mantener sus habilidades, o incluso para decidir qué habilidades deberían ser entrenadas (por ello, cuando los grupos diana son ancianos pueden recibir el nombre de evaluación de la plasticidad cognitiva).

Tabla 2

Concordancia del estatus de ganador entre dos pruebas de potencial de aprendizaje para un grupo de ancianos. Datos extraídos de Calero, Navarro, Arnedo, G<sup>a</sup>-Berben y Robles (2000). AVLT-PA= Auditory Verbal Learning Test- Potencial de Aprendizaje. WW-PA= Wisconsin Card Sorting Test- Potencial de Aprendizaje

N= 60	Con ganancia sig. en WW-PA.	No ganancia en WW-PA.
Con ganancias sig. en AVLT-PA.	24	17
No ganancia en AVLT-PA.	10	36

En todos estos casos se estudia, y parece ponerse de manifiesto, su validez y, lo que es más importante, su utilidad. Evidentemente, debemos de seguir trabajando y consolidando la

evaluación del potencial de aprendizaje, pero creemos que aún no hemos llegado al techo de sus posibilidades de aplicación.

## Referencias

- Baltes, P., Dittman-Kohli, F. y Kliegl, R. (1986). Reserve Capacity of the Elderly in Aging-sensitive Test of Fluid Intelligence: Replication and Extension. *Psychological and Aging*, 1(2), 172-177.
- Baltes, M., Kühl, K. y Sowarka, D. (1992). Testing the Limits of Cognitive Reserva Plasticity: A promising Strategy for Early Diagnosis of Dementia? *Journal of Gerontology*, 47, 165-177.
- Brown, A.L., Campione, J.C., Webber, L.S. y McGilly, K. (1992). Interactive Learning Environments. A New Look at Assessment and Instruction. En B.R. Guilford y M.C. O'Cooner (Eds.), *Changing Assessment. Alternative View of Aptitude, Achievement and Instruction*. Boston: Kluwer A.P. (pp. 121-212).
- Budoff, M., Meskin, J. y Harrison, R.H. (1971). Educational Test of the Learning Potential Hypothesis. *American Journal of Mental Deficiency*, 76, 159-169.
- Calero, M.D. (1995). *Modificación de la inteligencia. Sistemas de evaluación e intervención*. Madrid: Pirámide.
- Calero, M.D. y Márquez, J. (1998). Psychometric Properties of a Learning Potential Test for Reading: the Picture Word Game. *European Journal of Psychological Assessment*, 14(2), 124-133.
- Calero, M.D., Navarro, E., Arnedo, M.L., G<sup>a</sup>-Berben, T.M. y Robles, P. (2000). Estimación del potencial de rehabilitación en ancianos con y sin deterioro cognitivo asociado a demencias. *Revista Española de Geriatria y Gerontología*, 35(2), 44-50.
- Carlson J. y Wiedl, K.H. (1979). Towards a Differential Testing Approach: Testing-the-limits Employing the Raven Matrices. *Intelligence*, 3, 323-334.
- Carlson J. y Wiedl, K.H. (2000). The Validity of Dynamic Assessment. En C.S. Lidz y J.G. Elliot (Eds.), *Advances in Cognition and Educational Practice. Vol. 6*. Amsterdam: JAI (pp. 681-712).
- Cronbach, L.J. y Furby, L. (1970). How We Should Measure «Change» - Or Should We? *Psychological Bulletin*, 74(1), 68-80.
- Dwyer, C.A. (1996). Cut Scores and Testing: Statistics, Judgment, Truth, and Error. *Psychological Assessment*, 8(4), 360-362.
- Elliot, J.G. (2001). All Testing is Dynamic: A response to Sternberg and Grigorenko. *Issues in Education*, 7(2), 185-192
- Embretson, S. (1987). Toward Development of a Psychometric Approach. En C.S. Lidz (Ed.), *Dynamic Assessment. An interactional Approach to Evaluating Learning Potential*. N.York: The Guilford Press (pp. 107-136).
- Embretson, S.E. y Prenovost, L.K. (2000). Dynamic Cognitive Testing: What Kind of Information is Gained by Measuring Response Time and Modifiability? *Educational and Psychological Measurement*, 60(6), 837-863.
- Fernández-Ballesteros, R. y Calero, M.D. (1993). Measuring Learning Potential. *International Journal of Cognitive Education And Mediated Learning*, 3(1), 9-20.
- Fernández-Ballesteros, R. y Calero, M.D. (1995) Training Effects on Intelligence of Older Persons. *Archives of Gerontology and Geriatrics*, 20, 135-148.
- Fernández-Ballesteros, R., Calero M.D., Campllonch, J.M. y Belchí, J. (1987/90). *EPA: Evaluación del potencial de aprendizaje*. Madrid: Mepsa.
- Fernández-Ballesteros, R. y Calero, M.D. (2000). The assessment of Learning Potential: The EPA Instrument. En C.S. Lidz y J.G. Elliot (Eds.), *Advances in cognition and Educational Practice. Vol. 6*. Amsterdam: JAI (pp. 293-324).
- Fernández Fernández, S. (1990). Diagnóstico curricular y diferencias de aprendizaje. *Psicothema*, 2(1), 37-56.
- Feuerstein, R, Rand, Y. y Hoffman, M.B. (1979). *The Dynamic Assessment of Retarded Performers. The Learning Potential Assessment Device: Theory, Instrument and Techniques*. Baltimore: UPP.
- Gerber, M.M. (2001). All teachers are Dynamic Tests. *Issues in Education*, 7(2), 193-200.
- Guthke J. y Stein, H. (1996). Are Learning Tests the Better Version of Intelligence Tests? *European Journal of Psychological Assessment*, 12(1), 1-13.
- Haywood, H.C. (1993). As Mediational Teaching Style. *International Journal of Cognitive Education and Mediated Learning*, 3(1), 27-38.
- Haywood, H.C. (2001). What is Dynamic «Testing»? A response to Sternberg and Grigorenko. *Issues in Education*, 7(2), 201-210.
- Hessels, M.G.P. y Hamers, J.H.M. (1993). A Learning Potential Test of Ethnic Minorities. En J.H.M. Hamers, K., Sijtsma y A.J.J.M. Ruijsenaars (Eds.), *Learning Potential Assessment, Theoretical, Methodological and Practical Issues*. Amsterdam: Swets and Zitlinger (pp. 285-312).
- Kaniel, S. (2001). Not all Testing is Dynamic Testing: Overcoming the Fuzzy Identity of Dynamic Assessment. *Issues in Education*, 7(2), 211-224.
- Lauchlan, F. y Elliot, J. (2001). The Psychological Assessment of Learning Potential. *British Journal of Educational Psychology*, 71, 647-665.
- Lezak, M. (1995). *Neuropsychological Assessment*. N. York: Oxford U.P. (3<sup>a</sup> edición).
- Lidz, C.S. y Macrine, S.L. (2001). An alternative Approach to the Identification of Gifted Culturally and Linguistically Diverse Learners. The Contribution of Dynamic Assessment. *School Psychology International*, 22(1), 74-96.
- Lidz, C.S. y Thomas, C. (1987). The Preschool Learning Assessment Device: Extension Of A Static Approach. En C. Lidz (Ed.), *Dynamic Assessment: An Interactional Approach To Evaluating Learning Potential*. N. York, the Guilford Press (pp. 123-164).
- Lozano Fdez., L.M., G<sup>a</sup> Cueto, E. y Gallo Alvaro, P. (2000). Relación entre motivación y aprendizaje. *Psicothema*, 12(2), 344-347.
- May, K. y Nicewander, W.A. (1998). Measuring Chage conventionally and Actively. *Educational and Psychological Measurement*, 15(6), 882-897.
- Orrantía, J. Morán, M.C. y Gracia A.D. (1998). Evaluación estática vs. Evaluación dinámica. Una comparación experimental. *Estudios de Psicología*, 61, 35-50.
- Raven, J.C. (1969) *Progressive Matrices Set I and set II*. Londres: H.K. Lewis and co.
- Resing, W.C.M. (1997). Learning Potencial Assessment: The alternative for measuring intelligence? *Educational Child Psychology*, 14(4) ,68-82.
- Resing, W.C.M. (2001). Beyond Binet: All Testing Should be Dynamic Testing. *Issues in Education*, 7(2), 225-236.
- Robles, A. y Calero, M.D. (2003). Evaluación del potencial de aprendizaje en niños con síndrome down. *Siglo Cero*, 34(2), 14-25.
- Schöttke, H. Bartram, M. y Wiedl, K.H. (1993). Psychometric Implications of Learning Potential Assessment: A Typological Approach. En J.H.M. Hamers, K. Sijtsma y A.J.J.M. Ruijsenaars (Eds.), *Learning Potential Assessment, Theoretical, Methodological and Practical Issues*. Amsterdam, Swets and Zeitlinger (pp. 153-173).
- Sternberg, R. y Grigorenko, E.L. (2001). All testing is dynamic Testing. *Issues in Education*, 7(2), 137-170.
- Sternberg, R. y Grigorenko, E.L. (2002). *Dynamic Testing*. Cambridge: Cambridge U.P.
- Tzuriel, D. (2001). Dynamic Assessment is not Dynamic Testing. *Issues in Education*, 7(2), 237-250.
- Tzuriel, D. y Klein, P.S. (1987). Assessing the Young Child: children's Analogical Thinking Modifiability En C.S. Lidz (Eds.), *Dynamic Assessment*. N. York: the Gilford Press (pp. 268-282).
- Vygotski, L.S. (1978). *El desarrollo de los procesos psicológicos superiores*. Madrid, Grijalbo.
- Wiedl, K.H., Wienöbst, J. y Schöttke, H. (2000). Estimating rehabilitation potential in schizophrenic subjects. En H.D. Brenner, W, Böker y R. Genner (Eds.), *The treatment of schizophrenia: status and emerging trends*. Seattle: Hogrefe and Hubers Publishers (pp. 145-167).