University of Groningen

A review scrutinising the consequential validity of dynamic assessment

Tiekstra, Marlous; Minnaert, Alexander; Hessels, Marco G.P.

Published in:
Educational Psychology

DOI:
[10.1080/01443410.2014.915930](https://doi.org/10.1080/01443410.2014.915930)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2016

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Tiekstra, M., Minnaert, A., & Hessels, M. G. P. (2016). A review scrutinising the consequential validity of dynamic assessment. *Educational Psychology*, 36(1), 112-137.
<https://doi.org/10.1080/01443410.2014.915930>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Educational Psychology

An International Journal of Experimental Educational Psychology

ISSN: 0144-3410 (Print) 1469-5820 (Online) Journal homepage: <http://www.tandfonline.com/loi/cedp20>

A review scrutinising the consequential validity of dynamic assessment

Marlous Tiekstra, Alexander Minnaert & Marco G.P. Hessels

To cite this article: Marlous Tiekstra, Alexander Minnaert & Marco G.P. Hessels (2016) A review scrutinising the consequential validity of dynamic assessment, Educational Psychology, 36:1, 112-137, DOI: [10.1080/01443410.2014.915930](https://doi.org/10.1080/01443410.2014.915930)

To link to this article: <https://doi.org/10.1080/01443410.2014.915930>

Published online: 15 May 2014.

Submit your article to this journal [↗](#)

Article views: 635

View Crossmark data [↗](#)

Citing articles: 7 View citing articles [↗](#)

A review scrutinising the consequential validity of dynamic assessment

Marlous Tiekstra^{a*}, Alexander Minnaert^a and Marco G.P. Hessels^b

^a*Orthopedagogy & Clinical Educational Science, University of Groningen, Groningen, The Netherlands;* ^b*Educational Sciences, University of Geneva, Geneva, Switzerland*

(Received 19 July 2013; final version received 3 April 2014)

This literature review explored whether dynamic assessment procedures in psycho-educational practice might bridge the well-known gap between diagnosis and intervention. Due to a learning phase included in the testing procedure, qualitative information about the child's learning needs can be revealed by means of dynamic assessment. The question is, however, what the consequential validity, i.e. the extent to which assessment influences instructional and learning processes, of dynamic assessment procedures really is. The review of 31 articles that met the inclusion criteria showed that proximal consequential validity of dynamic assessment is warranted, but distal consequential validity is warranted to a lesser extent (e.g. some guidelines for practice). Furthermore, it can be noticed that motivational aspects never played an explicit role during learning phases. In order to design student-tailored interventions following dynamic assessment, there is a need for more explicitness of learning phases and types of feedback in the development of these instruments.

Keywords: dynamic assessment; psycho-educational intervention; consequential validity; learning problems

Introduction

Nowadays, in cohesion with current inclusive education policies, it is a challenge for educational practitioners to include children having severe learning difficulties into regular education (UNESCO, 1994). To be able to adopt these children, a proper assessment of the child's capacities is requested. On top of this, in case of special needs education, individualised education plans (IEP) are obligatory in order to construct adequate interventions. This, in turn, requires an indication of the child's ability to learn. Not only is the necessity for appropriate assessment procedures underlined here, but also the need for appropriate intervention guidelines. Nevertheless, in practice, the gap between diagnosing and intervention is far too often neglected or even disregarded (Minnaert & Vermunt, 2006; Resing, 2006; Ruijsenaars, 2001; Shapiro, 1987; Shapiro & Kratochwill, 2000; Wilson, 1998).

A way to resolve this issue is to highlight the importance of consequential validity during assessment procedures. Additionally, already in 1989, Messick pointed at the lack of focus on consequences of test use in validity research of tests. He stated that validity refers to the 'degree to which empirical evidence and theoretical

*Corresponding author. Email: M.Tiekstra@rug.nl

rationales support the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores' (Messick, 1989, p. 5). Accordingly, Gielen, Dochy, and Dierick (2003) investigated the influence of assessment on learning of students, which could be defined as one form of consequential validity of assessment. Besides, consequential validity refers to the influence of assessment on the instructional process (Messick, 1995). Furthermore, a division can be made in the extent to which or the moment that consequential validity occurs: proximal consequential validity refers to the consequences that influence learning and teaching during the testing procedure, while distal consequential validity denotes consequences of testing beyond the testing procedure, i.e. later on in the instructional processes in the classroom.

In static testing, both types of consequential validity can profoundly be questioned. Amongst others, Shapiro and Kratochwill (2000) pointed at the great demands from educational practice for cues following the outcomes of an IQ test. A dynamic assessment procedure is a more promising instrument to bridge the gap between assessment and instruction. During dynamic assessment a learning phase is included in which the examiner has an active role. During testing the examinee is being taught the prerequisites of the test procedure, and interaction plays a major role. In general, the aim of dynamic assessment is twofold, namely to be able to make a more accurate evaluation of a child's capacities and to provide useful information for clinical purposes. One could state that during dynamic testing the proximal consequential validity is warranted since an explicit opportunity to learn is provided during testing, and the examiner should act in such way that an optimisation of the child's learning is revealed. Moreover, to what extent the examinee profits from explanations and feedback is measured, which could give clues for future actions in the classroom. Therefore, dynamic testing could be a very promising tool to bridge the gap between diagnosis and intervention.

Dynamic assessment

The theoretical framework for dynamic assessment procedures is based on the theory of Vygotsky and his conceptualisation of the Zone of Proximal Development (ZPD) (Vygotsky, 1978). This ZPD is the difference between the actual independent achievement of the child and the performance level of the child when tailored help is provided (see also Lidz, 1995). Teaching in the ZPD helps fulfilling the child's learning potential. Hence, the aim of the learning phase in dynamic assessment procedures is to optimise learning processes by means of teaching learning strategies or to provide scaffolding and prompts during learning processes.

Due to teaching and consequently the active role of the examiner, the testing situation is adaptive to the examinee. In other words, during dynamic assessment the responsiveness of the child to adaptive teaching will be evaluated which provides an individualised diagnosis of the child's ability to learn and the learning processes that have been addressed. This is of value while diagnosing children, especially for children having learning difficulties. Resing (2006) underlined the fact that a dynamic approach provides qualitative information and much more resembles classroom learning than static procedures. Meijer (2001) demonstrated that dynamic assessment procedures are less biased than static measures according to test anxiety and the lack of self-confidence. Several authors showed the added value of dynamic procedures in the assessment of at-risk children (Beckmann, 2006; Carlson & Wiedl, 1992; Hessels, 1993, 2000; Resing, 1990; Schlatter & Büchel, 2000).

Different approaches to dynamic assessment exist, e.g. domain-specific vs. general capacity evaluation or focus on intervention vs. focus on evaluation. The main distinction in approaches, however, lies in diverging goals and, thus, manifests in two different procedures (distinction in role of examiner). In this, there is the clinical approach of Feuerstein (Feuerstein, Rand, & Hoffman, 1979) on the one side supported by, amongst others, Tzuriel (Shamir & Tzuriel, 2004), wherein it aims at the evaluation of the cognitive modifiability of the child during mediated learning experiences. On the other side, there are standardised approaches: dynamic tests or so-called learning potential tests, aiming at the evaluation of learning abilities of children. These tests try to measure to what extent a child is able to apply new information that has been taught during the learning phase. In a (pre-test –) training – post-test format the learning abilities of a child are assessed by measuring the extent to which the child has profited from the learning phase. In another format, the train-within-test format, the examinees are offered graduated prompting or feedback. Both formats are standardised and require an active role of the examiner in teaching strategies required for resolving test items.

Standardisation of testing procedures makes it possible to replicate testing and diminish the influence of personal characteristics of the examiner, which is a condition to meet scientific criteria for tests (objectivity, reliability and validity). Moreover, criteria for the identification of disabilities applied by the World Health Organisation or as indicated in the Diagnostic Statistic Manual are based on assessments showing satisfactory psychometric characteristics.

During dynamic assessment examiners play an important role, since it is their duty to optimise performances of the examinee. In dynamic tests this role is more standardised, whereas it is not in the clinical approaches. Dynamic testing is regarded as a promising tool to estimate the real capacities of at-risk children and to make better distinguishes between children with learning disorders and children having learning delays (as a consequence of their learning environments and former learning experiences). Dynamic tests fit into these requirements of psychometric characteristics. Therefore, the focus of this article is rather on dynamic testing as on the more clinical-based assessment approaches. More precisely, only dynamic tests with short-term learning phases are used for this review in order to make a clear distinction between tests that try to measure learning capacity as opposed to tests that try to change capacity. Consequently, in this article, the term dynamic assessment refers to a dynamic procedure with a standardised approach in which a short-term learning phase is included in testing.

Aim of the present review

Currently, articles about dynamic assessment merely focus on comparisons to traditional (static) measures or on psychometric characteristics. Since one of the aims of dynamic assessment is to optimise learning, the purpose of this review is to focus at the dynamic part in dynamic assessment and to examine what has been written about the consequential validity of dynamic tests. It is of added value to investigate what the consequences of dynamic assessment are for practice. Over time, quite a few dynamic assessment procedures have been developed. In order to clarify what makes dynamic assessment ‘dynamic’, the learning phases of dynamic assessment procedures will be scrutinised. The learning phases in these procedures differ latently and manifestly. The differences will be revealed due to a screening of the

presence of cognitive, metacognitive or motivational strategies (Dina & Efklides, 2009; Wigfield, Byrnes, & Eccles, 2006) as a goal of the examiner's activities during the learning phase. In other words, the focus is on the activities the examiner carries out and whether these are more cognitive-, regulative- or affective-oriented. Cognitive strategies are strategies that the student applies to process information, e.g. relating, structuralising, analysing, memorising and applying. Metacognitive strategies are strategies to regulate learning processes, like planning, evaluating, adjusting, diagnosing and reflecting. Whereas motivational strategies are more affective-oriented, like adjustment of attributions, values, emotions, expectations and concentration. A further elaboration on these strategies can be found in the classification schemes of Vermunt (1996).

Moreover, the outcomes of different dynamic instruments will be investigated to unravel the value of dynamic assessment for construction of interventions that might follow the assessment. In this way, the consequential validity of each instrument will be emphasised.

Over years, several dynamic assessment procedures have been developed for higher education (e.g. Antón, 2009; Meijer & Elshout, 2001; Nirmalakhandan, 2009). In these assessment procedures the focus is more on career or professional perspectives than on the optimisation of development of young individuals. This review, however, aims at more insight around the construction of adequate classroom interventions for children with some kind of special needs, since these will contribute to the optimisation of development of children. Moreover, in order to optimise a child's development (or in other words, to adapt the curriculum as best as possible to the child's learning needs), an early recognition of a child's learning difficulties is necessary. An early evaluation of a child's learning weaknesses and strengths is a prerequisite to provide early intervention, which, in turn, paves the way for optimal development. Therefore, this review will only include dynamic tests estimating the ability, capacity or potential to learn of children in regular primary (inclusive) educational settings and the transition to secondary or middle schools. Hence, studies carried out in grades one to eight in regular education are the review's objectives.

Method

The literature review consisted of an extensive computer search. For this purpose, the database of Web of Science was used. This database includes the Science Citation Index Expanded, the Arts and Humanities Citation Index and Social Sciences Citation Index (SSCI). The use of this database implies that the references found are from articles that are published in international peer-reviewed scientific journals. References from 1995 until 2011 have been included.

In order to capture as many relevant references as possible, a 'search model' had been developed. The model tries to combine several keywords into three categories which were regarded as most valuable to find literature about dynamic assessment. The three categories were the following: potential, dynamic and ability/capacity. The category *potential* enclosed the keywords 'learning potential' and 'students' potential'. About 245 articles had been found with these keywords. The second category *dynamic* comprised the keywords 'dynamic assessment'; 'dynamic test'; 'dynamic tests'; and 'dynamic testing'. This resulted in 2828 articles. The category *ability/capacity* enclosed the keywords 'cognitive abilities'; 'ability testing'; 'ability test(s)';

‘ability to learn’; and ‘capacity testing’. This third category had been used only in combination with one of the two other categories to narrow down the number of hits in the database. For example, the keyword ‘cognitive abilities’ was combined with ‘dynamic’ in a first search, and in a second search combined with ‘potential’. A total of 574 articles were found using these combinations.

The search strategy resulted in a total of 3420 articles after having deleted overlapping articles between (and within) the different categories. The list of 3420 references was screened in relevance to the field of learning and development. Subsequently, the total number of articles was reduced since they did not meet the inclusion criteria for the review study. In line with the previous section, the following inclusion criteria were formulated:

- (1) The article fully describes an empirical study focusing at a dynamic testing procedure to assess cognitive abilities.
- (2) The dynamic testing procedure described in the article is standardised and includes a short-term learning phase.
- (3) The study focuses on children aged 6–14 years (in the case of special education, a division has been made based on the average curriculum taught that fits these ages).
- (4) The study has been carried out in regular primary or special needs education.

The greatest exclusion of articles was caused by the first criterion, since a large part of the articles found did not describe a procedure to assess cognitive abilities. Moreover, some references were excluded since they were not full articles, such as book reviews, comments or abstracts for presentation at an international conference (45) or were not written in English (18). Finally, two references were untraceable: the articles could neither be retrieved from the internet, nor were available at any of the university libraries in The Netherlands, nor even after contact with the main author. The number of articles that has been used for the literature review is 31.

Results

Tables 1 and 2 present an overview of the results following the literature review. Table 1 summarises the empirical studies according to the instruments being used and their psychometric properties (if described). Table 2 zooms in on the dynamic tests that have been used in the studies. The aim of Table 2 is to reveal differences and similarities between the dynamic tests with respect to their learning phases.

It is emphasised, though, that psychometric properties described below should be considered critically. According to the authors, some methodological errors exist. Reliability measures, for example, vary between Cronbach’s alpha or interrater agreements, and some studies use test–retest reliability. The latter form does, however, not refer to reliability in this context, but rather to a measure for construct validity of the dynamic test. After all, test–retest reliability can be seen as a measure for stability, which is not in accordance with the basis of dynamic testing.

As can be seen in Table 1, slightly more than half (18 of 31) of the studies enclosed in the review have been carried out in North America (USA/Canada), the remaining in Europe, Israel or Africa. In line with our inclusion criteria, dynamic assessment procedures described were mostly developed for children in grades 1–8,

Table 1. Overview of articles selected for this review.

Authors (year of publication)	Country	N	Subjects	Dynamic test	Test procedure	Reliability	(predictive) validity	Consequential validity
Compton et al. (2010)	USA	355	Grade 1	DA pseudo words	Train within test	Test-retest (4 weeks): .72	Range: .11–.65	Scores indicate instructional scaffolding level needed to correctly solve items (min. 3; max. 15). The lower scores, the less scaffolding
de Beer (2010a)	South Africa	79 (boys)	mean age boys: 12.44	Learning Potential	Test-training-test	α .93–.98	School achievement: boys: $R^2 = .33$	Post-test score, difference score and composite score
		72 (girls)	mean age girls: 11.18	Computerised Adaptive Test (LPCAT)			girls: $R^2 = .32$	
Dolores Calero, et al. (2011)	Spain	127	7–11 y	Adapted subtests of the LPAD: Positions test	Train within test	α .70–.95	Indirectly taken into account by making use of Discriminant Analysis	Gain score, based on assistance needed
				The organiser test	Test-training-test			Raw scores on pre-test, post-test, and gain score
Elleman et al. (2011)	USA	100	Grade 2 mean age 8 y3 m	Stencil-design test DA inference generation	Test-training-test Test-training-test – transfer	α .76 Interrater agreement 97.8%	Cognitive test: 4% unique variance explained by DA	Gain score, based on assistance needed 3 scores: number of prompts needed, transfer score and total score
Fabio (2005)	Italy	287			Train within test	α .86	IQ: .48 (Raven CPM)	

(Continued)

Table 1. (Continued).

Authors (year of publication)	Country	N	Subjects	Dynamic test	Test proce- dure	Reliability	(predictive) validity	Consequential validity
Fuchs, Compton, Fuchs, Bouton, and Caffrey (2011)	USA	318	6–11 (mean age: 8 y 2m) Grade 1	Dynamic Test of Potential Intelligence DA pseudo words	Train within test	Test–retest (4 weeks): .72	School achievement: .27–.36 2.3% unique explained variance in word identification 1.0% unique explained variance in reading comprehension	General Modifiability Index (= score based on need for help) Scores indicate instructional scaffolding level needed to correctly solve items (min. 3; max. 15). The lower scores, the less scaffolding
Fuchs, Compton, Fuchs, Hollenbeck, Craddock, and Hamlett (2008)	USA	122	Grade 3	DA algebra skills	Train within test	Indirectly taken into account by making use of Structural Equation Modelling (SEM) The authors refer to the above- mentioned article of 2008	Math (school achievement) .41–.60	Scores indicate instructional scaffolding level needed to correctly solve items (min. 0; max. 21). The more scaffolding, the less points
Fuchs, Compton, Fuchs, Hollenbeck, Hamlett, and Seethaler (2011)	USA	122	Grade 3	DA algebra skills	Train within test		Improvement of classifications due to DA: 48 to 73%	Scores indicate instructional scaffolding level needed to correctly solve items (min. 0; max. 21). The more scaffolding, the less points

Gillam et al. (2011)	USA	64	6–8.5	Non-verbal DA phoneme deletion	Train within test	α .88	Construct val.: r .84 Word-level reading: r .54 ?	Scores reflecting number of prompts needed (min. 21; max 105) Static score (storylength and syntax) + qualitative interpretation of errors (audio-taped)
Gummersall and Strong (1999)	USA	30 (exp. 1) + 12 students with language problems (exp. 2)	8–9 (grade 3)	Narrative skill assessment	Test– training–test	Interrater agreement 99.2% to 100% .68 to .98		
Hamers et al. (1996)	The Netherlands	500	5 y 4 m–7 y 9 m	The Learning Potential Test for Ethnic Minorities (LEM)	Train within test	Range α .88–.92	IQ: .45–.68 School achievement: .32–.69	standardised LP scores (norm referenced)
Jeltova et al. (2011)	USA	1332	Grade 4	Group- administered dynamic assessment (GDA) mathematics	Test– training–test	IRT reliability .79–.84 for subtests	Taken into account by making use of HLM	Raw scores on pre- and post-test
Jitendra and Kameenui (1996)	USA	46	Grade 3 (8 y 9 m–9 y 10 m)	Mathematical word problems	Train within test	Interrater agreement 92 to 100%	?	Number of prompts and time elapsed
Kanevsky and Geake (2004)	Canada	25	8 y 10 m– 10 y 1 m	Mathematical pattern- recognition tasks	Train within test	Interrater agreement 92 to 96%	This study: IQ: r = .34	Need of support (total number of hints + qualitative observations)
Katz et al. (2007)	Israel	381	6–12 y	The Dynamic Occupational Therapy Cognitive Assessment for children (DOTCA-Ch)	Test– training–test	ICC (interrater) .87–.99 Int. consistency α .61–.77	Range .36–.71	Level of mediation needed: cut-off scores + dynamic percentage scores

(Continued)

Table 1. (Continued).

Authors (year of publication)	Country	N	Subjects	Dynamic test	Test pro- cedure	Reliability	(predictive) validity	Consequential validity
Kirkwood et al. (2001)	USA	202 (heterogeneous learning problems)	7–11	Rey-Osterrieth Complex Figure (ROCF) – dynamic adaptation	Test– training–test	?	?	Raw scores based on Developmental Scoring System are divided in levels; division in two subgroups (benefit/no benefit)
Larsen and Nippold (2007)	USA	50	Grade 6 Mean age 12.2 (age range: 10 y 9 m–12 y 10 m)	Dynamic Assessment Task of Morphological Analysis (DATMA)	Train within test	Interscorer agreement 95 to 100% Internal consistency is strong*	DATMA & PPVT-III: $r=.36$ School achievement: DATMA & OSA: $r=.50$	Scores based on need for prompts: raw scores, max. 75
Lauchlan and Elliott (2001)	UK	30 (special education)	Mean age 9; moderate/ severe learning difficulties	The Children's Analogical Thinking Modifiability test (CATM)	Preliminary phase–test – training–test	Rel. proven*	Val. proven*	Raw scores (min. 0; max. 18)
Lidz and Macrine (2001)	USA	81	Grade 1–5	Naglieri Non- verbal Ability Test – modified for dynamic administration (NNAT/DA)	Test– training–test	?	Concurrent validity K- ABC .64–.74	Standard scores, based on NNAT norms

Peltenburg et al. (2009)	The Netherlands	37 (special education)	8.9–12.11 (mean age 10.5 years) Average curriculum redesigned for 7–8 years	CITO Monitoring Test for Mathematics, redesigned for ICT and DA	Train within test	?	Raw scores + qualitative observations on strategy use
Peltenburg et al. (2010)	The Netherlands	37 + 43 (special education)	8–12 Average curriculum level: end grade 2	CITO Monitoring Test for Mathematics, redesigned for ICT and DA	Train within test	Indirectly taken into account by making use of differential item functioning (DIF)	Raw scores + qualitative observations on strategy use
Resing et al. (2009)	The Netherlands	109 (55 ethnic-minority children)	7–9y	Seria-Think – adapted version	Test–training–test	Pre-post-test correlation: .57; .61	Number of insertions and measurements + strategy-level categories
Sternberg et al. (2002)	Tanzania	358 (+100 children in control group)	11–13 y	Syllogisms	Test–training–test	Cognitive tests: .21–.31 School achievement: .29–.33	Number of correct answers in pre-test, training and post-test
				Sorting	Test–training–test	Cognitive tests: .29–.35/.32–.38 School achievement: .35 to .42/.34 to .50	Number of perseverative errors and categories produced + number of hints (training phase)
				Twenty Questions	Test – training – test	Test – retest reliability: .66	Ratio of constraint-seeking questions to total number of questions

(Continued)

Table 1. (Continued).

Authors (year of publication)	Country	N	Subjects	Dynamic test	Test proce- dure	Reliability	(predictive) validity	Consequential validity
Swanson (1995a)	USA	Exp. 1: 61 Exp. 2: 191	Mean age 10.63 Mean age 10.66	S-CPT (Swanson- Cognitive Processing Test)	Train within test	α' s: .80–.95 Overall reliability: α . 98	achievement: .15 to .28 14 to 26% explained variance on school achievement 32% explained variance in reading recognition 40% explained variance in reading growth	In addition to initial standard score: gain score, probe score and maintenance score
Swanson (1995b)	USA/ Canada	1611 (10% special educ.)	Age range: 4.5–78.6	S-CPT	Train within test	α 's: .82–.95		7 composite standard scores ($M=100$; SD $=15$)
Swanson (2011)	USA	78	Mean age 11.62	4 subtests of S- CPT: rhyming, digit-sentence, visual matrix, mapping and directions	Train within test	Subtests correlation range .40–.80		Composite standard scores. Initial score, gain score, probe score and maintenance score
Swanson and Howard (2005)	USA	70 students with reading (and math) problems	Mean age 11.93	2 subtests of S- CPT: rhyming and digit- sentence	Train within test	α 's .76–.85	25% extra (unique) variance on top of IQ	Initial score, gain score, probe score and maintenance score
Tiekstra et al. (2009)	Switzerland	46 (special education)	15 y 6 m– 19 y 4 m (mean age 17 years) Average curriculum 8–10 years	Hessels Analogical Reasoning Test (HART)	Training – post-test	α 's .69–.92 This study: .65	IQ: –.18 School achievement (CLT): .57 40% prediction in CLT on top of IQ	Raw scores (min. 0; max. 20)

Tzuriel (2000)	Israel	48	Grade 2 (mean age = 7.5 y)	Serial-Think Instrument	Preliminary phase – test– training–test	α 's .37–.85	School achievement (math): $R^2 = .19$ $R^2 = .54$ Val. ok*	Raw scores, divided in performance level (max. 20); number of insertions; number of measurements
Tzuriel and Kaufman (1999)	Israel	52	Grade 1 (mean age = 7.16 and Ethiopian group: 6 y – 7 y 6 m)	The Children's Analogical Thinking Modifiability test (CATM) The Children's Inferential Thinking Modifiability test (CITM)	Preliminary phase – test– training–test Test– training–test – transfer phase	Rel. ok* α ' s .82–.90	Val. ok*	Raw scores: 'all-or- none' and 'partial credit'
Viljoen et al. (2007)	USA	190 (subgroup)	11–13y	MacArthur Competence Assessment Tool–Criminal Adjudication (MacCAT-CA)	Train within test	Intraclass correlation .64–.91	?	Raw scores

Notes:

*Only information indicated in the article.

?Lacking information in the article.

but the age ranged 4,5 years to 78,6. Some articles report studies that have been carried out in special education. In these studies the mean chronological age goes beyond the age of our interest, however, the average curriculum taught or tested fits our criterion age range. Roughly half of the articles found (18 of 31) describe studies with a sample size larger than or equal to 100.

Test procedures differed from train-within-test format to (pre-test –) training – post-test format. Overall, reliability, as mentioned in the method section as well as in results section in the articles, of the instruments was satisfactory, but a huge variation in reliability coefficients was encountered. The internal consistencies ranged from α .37 to .98, test–retest or pre-test – post-test reliabilities from .54 to .72 and interrater agreements from .64 – 1.00. Validity measures include different values and calculations, as is shown in Table 1. Predictive validity (if indicated in the articles) was of particular interest, since this form of validity matches one of the reasons to do an assessment in educational settings; namely, to provide an estimation or prediction of students' future learning outcomes. Therefore, a division was made of predictive validity based on comparisons of dynamic assessment scores with either school achievement test or static intelligence test (IQ) scores. The correlations with school achievement test scores varied from .11 to .71, and described explained variances from 1 to 54%. One study showed an explained variance of 40% in reading growth. Three studies demonstrate 4, 25 and 40% explained variance on top of IQ. The range of correlations with IQ tests varies from –.18 to .74. However, using an interpretation of the correlation between IQ and DA measure solely is too simplistic. A criterion measure is needed to be able to estimate the predictive validity.

Consequential validity

With regard to consequential validity of dynamic assessment procedures included in the review, it is of importance to investigate the possible consequences of these testing procedures in practice. During analysis of articles this was reflected by the extent to which the outcomes of dynamic assessment procedures were made explicit for practice. Indeed, the procedures differed in their extent of explicit information for practice. There were three categories in which the dynamic assessment procedures appeared to be divided. It is highlighted to what extent proximal consequential validity (raw or self-evident scores) and distal consequential validity (explicit information for practice) occur in assessment procedures in each of these categories.

Raw or standardised scores

Dynamic assessment procedures that result in a raw or standardised score have a minimal form of consequential validity. Although the LPCAT (de Beer, 2010a) provides an indication of the education level of the examinee compared to national education levels (de Beer, 2010b), the test does not result in explicit clues for the development of an intervention. Even so, in the studies with the LEM (Hamers, Hessels, & Pennings 1996), the NNAT/DA (Lidz & Macrine, 2001) and the study of Dolores-Calero, Belen, and Auxiliadora Robles (2011), the test scores are norm referenced, but what these scores are made of is not indicated in this article and thus, it is unclear whether the scores provide indications for practice. Other dynamic assessment instruments (Jeltova et al., 2011; Tiekstra, Hessels, & Minnaert, 2009; Tzuriel, 2000; Tzuriel & Kaufman, 1999; Viljoen, Odgers, Grisso, & Tillbrook, 2007) result

in raw scores. Lauchlan and Elliott (2001) describe a study that investigated the extent to which scores of the CATM could predict the learning progress of children following a cognitive intervention programme. Although this indicates some potential for distal consequential validity, the CATM scores, however, were only raw scores similar to the version applied in study of Tzuriel and Kaufman (1999) mentioned before.

Since a learning opportunity is provided during the above-mentioned procedures, the proximal consequential validity is warranted. The outcomes are represented by a number, and this number should be appropriate for predicting future learning outcomes. However, no explicit information about the learning processes is reflected in this number. The gap between diagnosis and intervention will be apparent after administration of these tests. To conclude, the distal consequential validity is not guaranteed according to these test outcomes.

Raw scores + indication of level of help needed

Several dynamic assessment procedures resulted in an indication of the level of support needed in addition to raw scores. Some of the procedures still result in an isolated number, which needs to be interpreted in a specific manner. Compton et al. (2010) and Fuchs, Compton, Fuchs, Bouton, and Caffrey (2011), for instance, demonstrated that a dynamic assessment procedure on pseudo words could forecast a child's responsiveness to intervention: the procedure provides a score which indicates the instructional scaffolding level needed which could link assessment to educational practices, nonetheless implicitly. In the studies of Fuchs et al. (2008, 2011) and Larsen and Nippold (2007) administration of scores is done in a similar way. Minimum and maximum scores are composed of points for items solved correctly, reduced (or increased) by the number of hints needed. The tests of Sternberg et al. (2002) and Elleman, Compton, Fuchs, Fuchs, & Bouton (2011) result in raw scores combined with the number of errors made or the number of hints required during training. This can be seen as a slightly more advanced form of proximal consequential validity than the previous category, but still, distal consequential validity is not warranted. No explicit clues for practice could be detected in these procedures.

This also holds for the instruments that result in multiple scores, like the S-CPT (Swanson, 1995a, 1995b, 2011; Swanson & Howard, 2005), the dynamic administration of the Rey-Osterrieth Complex Figure Test (Kirkwood, Weiler, Bernstein, Forbes, & Waber, 2001) and the adapted version of the Seria-Think Instrument (Resing, Tunteler, de Jong, & Bosma, 2009). In the latter, though, an estimation was made of the level of strategies used (least sophisticated strategy to advanced strategy) in addition to the raw scores. The link to instruction or intervention after the test is not made explicitly, and the estimation of the child's competences is central to these instruments.

Other procedures slightly touched distal consequential validity, since these added qualitative information to the raw scores. Gummertsall and Strong (1999) tested a dynamic procedure for narrative skill assessment. After completion of this assessment not only static scores are presented, but the qualitative interpretation of errors (audio-taped) provides useful individualised clues for intervention. Similarly, next to raw scores, Peltenburg, van den Heuvel-Panhuizen, and Doig (2009) and Peltenburg, van den Heuvel-Panhuizen, and Robitzsch (2010) add qualitative information about strategy use based on video observations which were made during the assessment

procedure. How this qualitative information is given back to the educational psychologists in order to set up an IEP remains, however, unclear. Therefore, these instruments only slightly address distal consequential validity.

Level of mediation needed/need of support

Some dynamic assessment procedures did not result in raw scores at all. These procedures have been grouped in this third category. Initially, one would think that this category provides the most advanced form of consequential validity since scores only reflect the level of help needed. It appeared, however, that these procedures also never explicitly mentioned clues for practice. For example, the instruments of Fabio (2005), Gillam, Fargo, Foley, and Olszewski (2011), Jitendra and Kameenui (1996), and Katz, Golstand, Bar-Ilan, and Parush (2007) resulted in scores composed of number of prompts needed during testing. The scores as such do not provide information for designing interventions. Yet, a qualitative analysis of the probes used by each child, however, can, since the standardised probes depended on the errors made. In this way, the scores provide more information for interventions than the outcomes of dynamic assessment procedures mentioned previously, although information remains implicit. Scores that provide more explicit information for practice can be found in the study of Kanevsky and Geake (2004). They investigated the interactions during testing: concrete, cognitive or metacognitive instruction by examiner and content-related or spontaneous (prompting) responses of student. The outcomes of the dynamic procedure indicate the need for support, composed of the total number of hints and time needed to complete the tasks, and added by qualitative observations.

To unravel the possibility of distal consequential validity of dynamic assessment procedures more into detail, learning phases should be scrutinised. Since the learning phase enhances a teaching component, which could provide valuable information for future classroom or intervention actions, the next paragraph describes the learning phases of each instrument (29 in total).

Learning phase

Table 2 shows differences and similarities between the dynamic tests with respect to their learning phases. Dynamic assessment procedures cover general or specific domains. Half of the tests try to assess the general ability of the child and the other half the abilities in specific domains as reading or mathematics. The HART (Tiekstra et al., 2009) and the DA inference generation (Jeltova et al., 2011) are the only instruments that have been group administered, including a group-wise training. Although, the LPCAT (de Beer, 2010a) and the test of Peltenburg et al. (2009, 2010) can be administered in groups too, since these are computerised tests.

Furthermore, it is of importance to scrutinise the learning phases of each dynamic assessment procedure in order to investigate the potential consequential validity for practice. Table 2 zooms in at the different learning phases with regard to their manifestations and contents.

Overall, one could state that a reasonable amount of instruments use the ‘graduated prompts’ approach. According to Resing et al. (2009), this implies a standardised training procedure in which prompts are prescribed and hierarchically ordered. The latter makes the testing procedure adaptive, since the examiner does not provide

Table 2. Overview of instruments used in the studies selected for this review.

Test	Domain (spec./general)	Administration (ind./group)	Test duration	Manifestation of learning phase	Content of learning phase		
					C	MC	M
DA pseudo words	Specific: reading	Individually	20–30 min	Graduated prompting (instructional scaffolding)	x	–	–
LPCAT	General	Individually & group wise (computerised)	?	Extra practice; general strategies and hints; feedback	x	*	–
Positions Test	General (visuo-spatial memory)	Individually	3 sessions of 50 min in total	Graduated instruction on failed items	x	–	–
The Organiser Test	General	Individually		Instruction on strategies, memory and organisation	–	x	–
Stencil-Design Test	General (perceptual structuration)			Training on visual transport and internal transformation of the stimulus	x	–	–
DA inference generation	Specific: reading	Individually	25 min–1 h	Graduated prompting + instruction	x	–	–
Dynamic Test of Potential Intelligence	General	Individually	14 items	Graduated prompting	x	–	–
DA algebra skills	Specific: mathematics	Individually	30–45 min	Graduated prompting (instructional scaffolding)	x	–	–
Non-verbal DA phoneme deletion	Specific: language	Individually	21 items	Systematic prompting + corrective feedback	x	*	–
Narrative skill assessment	Specific: language	Individually	?	Extra instruction (sentence-imitation task + picture cues); general feedback	x	*	–
LEM	General	Individually	?	Repeating items; non-verbal feedback; demonstration (graduated prompting)	x	*	–
GDA mathematics	Specific: mathematics	Group wise	50–70 min.		x	x	–

(Continued)

Table 2. (Continued).

Test	Domain (spec./general)	Administration (ind./group)	Test duration	Manifestation of learning phase	Content of learning phase		
					C	MC	M
Mathematical word problems	Specific: mathematics	Individually	?	Group-wise discussion and instruction on problem-solving strategies and thinking skills			
Mathematical pattern-recognition tasks	Specific: mathematics	Individually	14–56 min	Graduated prompting, feedback	x	x	–
DOTCA-Ch	General	Individually		Graduated support	x	x	–
ROCF – dynamic adaptation	General	Individually	Up to 1.5 h.	Prompts; feedback during test	–	–	–
DATMA	Specific: language	Individually	?	Prompting: structuralisation of figure	*	x	–
NNAT/DA	General	Individually	15 items	Graduated prompting	x	–	–
CITO monitoring	Specific: mathematics	Individually	15–20 min	Graduated prompting	x	–	–
Test for Mathematics – ICT/DA	General: inductive reasoning	Individually	3 sessions of 15–20 min (of which one session is training)	Visualisation and opportunity to use external memory	x	*	–
Serial-Think – adapted version	Specific: mathematics	Individually		Graduated prompting	x	x	–
Syllogisms	General	Individually	Total administration time range: 42–101 min. (mean time: 71 min.)	Feedback after pre-test, than teaching the skills to solve items.	x	–	–
Sorting	General (executive functioning)	Individually	Intervention time range: 11–28 min (mean time: 17 min)	Teaching skills to solve the task (visualisation)	x	–	–
Twenty questions	General (executive problem-solving skills)	Individually		Teaching skills to solve the task	x	–	–

S-CTP	General (working memory ability)	Individually	±150 min/±50 min. (abbreviated version: 5/6 subtests)	Graduated prompting	x	–	–
HART	General (analogical reasoning)	Group wise	20 min. training + 20 items post-test	Training on skills needed for solving items (prompting, extra instruction, verbalisation and visualisation)	x	x	–
Serial-Think Instrument	Specific: mathematics	Individually	Teaching phase: 30–40 min.	Instruction on strategies + feedback	x	x	x
CATM	General (analogical thinking)	Individually	Teaching phase : 30–40 min.	Instruction on problem-solving strategies	x	*	x
CITM	General (inferential thinking)	Individually	12 items each phase and 10 items transfer phase	Extra instruction + instruction on strategies	x	*	–
MacCAT-CA	Specific: adjudicative competences (legal understanding)	Individually	Only 6 out of 8 items of one subscale were dynamically administered	Extra instruction on failed items	x	–	–

Notes: Where x means explicitly mentioned in the article, * means implicitly mentioned in the article, and – means not mentioned at all in the article. C = cognitive learning strategies, MC = metacognitive learning strategies, and M = motivational learning strategies. ? = lacking information in the article.

more hints or prompts when the child has found the right answer. Other learning phases are composed of extra practice, extra instruction on failed items or on strategies and skills required to solve the tasks. During some of the learning phases general feedback is given.

In order to scrutinise the learning phases more carefully, their content, i.e. the aims of examiner's activities, has been examined. Therefore, a division in cognitive, metacognitive and motivational content has been made. The majority of tests contains a learning phase in which the examiner's activities are oriented towards cognitive strategies only. In a few instruments, metacognitive strategies are applied next to cognitive strategies. These are, for example, activities explicitly aimed at planning behaviour or evaluating one's own activities and not primarily focused on the solution of task.

In the instruments CATM and Seria-Think (Lauchlan & Elliott, 2001; Tzuriel, 2000; Tzuriel & Kaufman, 1999) a preliminary phase is included to familiarise the child with the testing procedure and to observe potential difficulties (Tzuriel, 2000). Lauchlan and Elliott (2001) report that the examiner addresses cognitive and affective factors during the learning phase of the CATM. In the article of Tzuriel (2000) about the Seria-Think Instrument, it is mentioned explicitly that during mediation (learning phase) the examiner gives feedback in order to assure the child's feeling about competence, even if the answer is not correct. This can be interpreted as an affective, or motivational, strategy of the examiner. The remaining instruments do not describe motivational factors other than some encouragements (e.g. Jeltova et al., 2011) which are not equivalent to motivational factors as described in the introduction.

It must be mentioned, though, that these results are based on rather vague and implicit information, in such way that most of the studies do not describe the metacognitive or motivational aspects explicitly. Mostly, only cognitive aspects were mentioned unambiguously.

Discussion

In an attempt to bridge the gap between diagnosis and intervention, which is nowadays a reasonable quest in educational practice, this study examined the contribution of dynamic assessment to this issue. Dynamic assessment is meant to be a promising tool for estimating capacities of at-risk children and for distinguishing appropriately between children with learning disorders and children having learning delays (as a consequence of their learning environments and former learning experiences). To date, this is still one of the challenges of educational practice. In spite of the advantages of dynamic testing procedures they are, still, too rarely employed in practice due to the robustness (in research as well as in practice) of traditional IQs. However, a lot of dynamic assessment procedures have been developed, demonstrating valuable psychometric properties and proving to be of added value in the assessment of capacities of at-risk children compared to static intelligence measures. Overall, they show a high predictive validity, and four studies (Elleman et al., 2011; Swanson, 2011; Swanson & Howard, 2005; Tiekstra et al., 2009) even demonstrate significant added predictive validity on top of IQ.

Over the last two years the contribution of articles about dynamic assessment to scientific journals has been increasing: more than one-third of the articles found for this review were published in 2010 and 2011. This trend also sustains in 2013, but

was beyond the scope of this article. The increased interest in dynamic assessment makes the necessity for review studies on this subject evident.

Since dynamic assessment implies a teaching component during testing, the question arose whether these dynamic assessment procedures provide valuable information for the construction of interventions following the test, and could, thus, bridge the gap between diagnosis and intervention. The outcomes of 29 dynamic instruments were examined with regard to their consequential validity and presented at a continuous range, diverging from proximal consequential validity to distal consequential validity. This means that outcomes vary from only raw scores to more qualitative information. Although one could state that qualitative observations during testing do provide a lot of information for intervention, it appeared that none of the instruments resulted in *explicit* clues for practice. Next to raw or standardised scores, there are tests that provide an indication of the level of help the child needs. This can be seen as a more advanced type of consequential validity, since it provides more information about the capacities of the child. Nevertheless, explicit clues for practice are still unknown.

This outcome of the review study underlines the fact that, although dynamic assessment procedures have potential, these procedures do not contribute to bridging the gap between diagnosis and intervention in their current formats. The remaining quest for bridging effects of dynamic assessment was also highlighted by Haywood (2012). Dynamic assessment procedures that result in raw scores only do contribute to proximal consequential validity, since an opportunity to learn is provided during testing. However, in these cases, the tests will particularly be used to better estimate learning capacities as opposed to traditional IQs, and not as indicative for the design of interventions or adaptations in the instructional process afterwards. Hence, distal consequential validity is not warranted in current dynamic assessment procedures. It should be mentioned, though, that it is likely that some of the articles describe instruments that have been developed for research purposes only, and not (yet) for psycho-educational practices. And, thus, do not describe consequences for practice explicitly. This aspect could have caused some biased results.

Despite the present absence of distal consequential validity in procedures, attempts to bridging effects are being made sporadically resulting in some qualitative information for practice. However, the move to fully distal consequential valid procedures would not be too far, and is actually near. The teaching component during testing provides the opportunity to carefully administering the actions that have been taken during learning phases. The analysis of actions of both examiner and examinee could be divided in, for example, different categories (i.e. cognitive, metacognitive and motivational). By explicitly mentioning these actions during the test administration, outcomes will provide information for practice. Outcomes of the testing procedures, then, not only indicate a score of learning potential (or raw score), but also enhance systematic cues for setting up IEPs.

Therefore, in addition to outcomes, learning phases of each dynamic assessment procedure have been scrutinised in this review. The content, i.e. the aims of examiner's activities, has been divided in cognitive, metacognitive and motivational strategies. Unfortunately, it turned out that this information was barely explicitly mentioned in the articles. Nonetheless, a striking conclusion can be made: most of the tests apply cognitive and/or metacognitive strategies during the learning phase, but motivational factors never played a major role explicitly. As has been noted before by Borkowski, Carothers, Howard, Schatz, and Farris (2007), self-regulation

strategies should be integrated in dynamic assessment. This review shows that progress still has to be made in this area. Only recently, Teo and Roodenburg (2013) demonstrated the additive effects of including self-regulation strategies during an assessment procedure. Moreover, in some dynamic tests general feedback is given (de Beer, 2010a; Gillam et al., 2011; Gummertsall & Strong, 1999; Hamers et al., 1996; Jeltova et al., 2011; Sternberg et al., 2002; Tzuriel, 2000), but the kind and type of feedback is not adequately mentioned or registered. The extent to which examinees benefit from feedback is an important factor in dynamic assessment. Woide, Beckmann, Elliott and Guthke (2005) demonstrated that motivational, affective and non-intellective factors influence the effects of feedback and that feedback increases the capacity to evaluate oneself. Although one of the promising aspects of dynamic assessment approaches is the attention for non-intellective factors (Lidz & Macrine, 2001), this study demonstrated that there is no explicit focus on these factors during learning phases. This is an interesting finding, since it is generally known that motivational factors interact with the learning process (e.g. Ryba, 1998) and test results (Meijer, 2001). Moreover, one of the groundings of dynamic assessment is that during testing social interaction takes place in order to provide a more motivating testing situation, which is less biased to e.g. test anxiety.

These aspects should, thus, play a role during learning phases. The type of feedback that has been given should be made explicit in the test result which will provide useful information for the instructional design in the classroom or for the design of individualised interventions. Motivating strategies of the examiner, like effort stimulation or explicit feedback on the performance of the task, have been demonstrated to result in different impacts on aspects of learning, such as task interest and self-efficacy, respectively (Perry, Turner, & Meyer, 2006; Prince, Minnaert, & Opdenakker, 2014).

These suggestions also underline the interdependency of cognitive, metacognitive and motivational strategies during learning processes. These aspects are indispensably related to each other, and thus, each of these aspects should play a prominent role during test administration, and, consequentially in the test result. Only in this way, an ecologically valid contribution to the consequential validity of dynamic assessment procedures can be made.

Of course, this relates to questions about the test duration, domain specificity and age group of the tests. Each of these variables has its implications for the question of consequential validity. However, the overall lack of motivational components and explicitness of learning phases underlines the importance of qualitative reports about the interplay between examiner and examinee during test administration, in addition to raw scores if applicable. This implies that examiners should be reflective by nature on both their own activities and the activities of the examinee.

There are, however, some important limitations while interpreting the outcomes of this review. Firstly, the literature search has been carried out in a database (Web of Science) which only includes peer-reviewed journal articles. Consequently, articles which have been published in non-SSCI journals are not warranted, as well as book chapters and dissertations. This implies a certain restriction to the scope of articles included in this review.

Secondly, there is no consensus about the definition of dynamic assessment in the field. Moreover, different approaches exist (standardised vs. clinical), which may cause confusion in research as well as in practice. Next to the robustness of traditional tests, this could contribute to the suppression of dynamic procedures in

practice. Therefore, it is suggested to use clear distinctions between dynamic assessment and intervention (see also Hessels-Schlatter & Hessels, 2009). In addition to these authors, we would like to maintain these distinctions to promote a clearer concept of dynamic assessment. Subsequently, in this article, the term dynamic assessment has been defined as a dynamic procedure with a standardised approach in which a learning phase is included in testing. This form of dynamic assessment is comparable to static testing. To our opinion, dynamic assessment does not contain an intervention phase expanded to multiple sessions during several weeks. In the latter, one is testing the effectiveness of an intervention, instead of making an evaluation of a child's capacities.

Conclusion

To conclude, dynamic assessment remains a promising tool to bridge the gap between diagnosis and intervention, but there is still room for improvement. Due to a learning phase included during the testing procedure, qualitative information can be revealed. This, in turn, provides an estimation and evaluation of the types of information that are essential to the child in order to complete a task. Eventually, this could contribute to the development of an adaptive intervention, which will provoke better learning results of the child.

Implications for practice

In order to establish this, however, the focus on consequences of assessment for educational practice should hold a more prominent place in the development of dynamic assessment procedures in future. There is a need for more explicitness of learning phases and types of feedback, in order to construct student-tailored interventions. Moreover, motivational factors should be highlighted during learning phases. Students encountering learning difficulties generally have experienced a lot of negative emotions and failure, and dynamic assessment procedures are more promising if they could anticipate on these factors pre-eminently. If the learning phase is administered and documented carefully, the consequential validity will be more warranted resulting in explicit clues for practice.

Additionally, the authors would like to make a plea for a more prominent place of consequential validity in test evaluation reports. When evaluating the quality of tests, it is important to focus on consequences for practice, next to internal validity criteria.

From a researcher's point of view, future research should focus on the bridging effect of dynamic assessment procedures to intervention practices. This could be done, e.g. by analysing qualitative (video) observations and focusing on motivational components during assessment procedures and consequences for educational practice. While investigating the distal consequential validity of dynamic assessment procedures, research contributes to practice, which is an important aim of science.

If the gap between diagnosis and intervention can be bridged, the quality of psycho-educational practices will be elevated. Since a careful administration of the learning phase provides useful information for interventions or instructional processes in the classroom, professional actions will be more grounded on empirical results. Educational psychologists play a prominent role in this process, being the core communicator between test result and interventions. If tests provide helpful

clues, next to raw scores, the quality of interventions and instructions will increase. Only in this way the learning outcomes of students will be optimised.

References

- Antón, M. (2009). Dynamic assessment of advanced second language learners. *Foreign Language Annals*, 42, 576–598.
- Beckmann, J. F. (2006). Superiority: Always and everywhere? On some misconceptions in the validation of dynamic testing. *Educational and Child Psychology*, 23, 35–49.
- Borkowski, J. G., Carothers, S. S., Howard, K., Schatz, J., & Farris, J. R. (2007). Intellectual assessment and intellectual disability. In J. W. Jacobson, J. A. Mulick, J. Rojahn, J. W. Jacobson, J. A. Mulick, & J. Rojahn (Eds.), *Handbook of intellectual and developmental disabilities* (pp. 261–277). New York, NY: Springer. doi:[10.1007/0-387-32931-5_14](https://doi.org/10.1007/0-387-32931-5_14)
- Carlson, J. S., & Wiedl, K. H. (1992). Principles of dynamic assessment: The application of a specific model. *Learning and Individual Differences*, 4, 153–166. doi:[10.1016/1041-6080\(92\)90011-3](https://doi.org/10.1016/1041-6080(92)90011-3)
- Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Gilbert, J. K., Barquero, L. A., ... Crouch, R. C. (2010). Selecting at-risk first-grade readers for early intervention: Eliminating false positives and exploring the promise of a two-stage gated screening process. *Journal of Educational Psychology*, 102, 327–340. doi:[10.1037/a0018448](https://doi.org/10.1037/a0018448)
- de Beer, M. (2010a). Longitudinal predictive validity of a learning potential test. *Journal of Psychology in Africa*, 20, 225–231.
- de Beer, M. (2010b). A modern assessment psychometric approach to dynamic assessment. *Journal of Psychology in Africa*, 20, 241–246.
- Dina, F., & Efklides, A. (2009). Metacognitive experiences as the link between situational characteristics, motivation, and affect in self-regulated learning. In M. Wosnitza, S. Karabenick, A. Efklides, & P. Nenniger (Eds.), *Contemporary motivation research: From global to local perspectives* (pp. 117–146). Cambridge, MA: Hogrefe & Huber.
- Dolores Calero, M., Belen, G. M., & Auxiliadora Robles, M. (2011). Learning potential in high IQ children: The contribution of dynamic assessment to the identification of gifted children RID C-7284-2011. *Learning and Individual Differences*, 21, 176–181. doi:[10.1016/j.lindif.2010.11.025](https://doi.org/10.1016/j.lindif.2010.11.025)
- Elleman, A. M., Compton, D. L., Fuchs, D., Fuchs, L. S., & Bouton, B. (2011). Exploring dynamic assessment as a means of identifying children at risk of developing comprehension difficulties. *Journal of Learning Disabilities*, 44, 348–357. doi:[10.1177/0022219411407865](https://doi.org/10.1177/0022219411407865)
- Fabio, R. A. (2005). Dynamic assessment of intelligence is a better reply to adaptive behavior and cognitive plasticity. *The Journal of General Psychology*, 132, 41–64.
- Feuerstein, R., Rand, Y., & Hoffman, M. B. (1979). *The dynamic assessment of retarded performers. The learning potential assessment device: Theory, instruments, and techniques*. Jerusalem: ICELP Press.
- Fuchs, D., Compton, D. L., Fuchs, L. S., Bouton, B., & Caffrey, E. (2011). The construct and predictive validity of a dynamic assessment of young children learning to read: Implications for RTI frameworks. *Journal of Learning Disabilities*, 44, 339–347. doi:[10.1177/0022219411407864](https://doi.org/10.1177/0022219411407864)
- Fuchs, L. S., Compton, D. L., Fuchs, D., Hollenbeck, K. N., Craddock, C. F., & Hamlett, C. L. (2008). Dynamic assessment of algebraic learning in predicting third graders' development of mathematical problem solving. *Journal of Educational Psychology*, 100, 829–850.
- Fuchs, L. S., Compton, D. L., Fuchs, D., Hollenbeck, K. N., Hamlett, C. L., & Seethaler, P. M. (2011). Two-stage screening for math problem-solving difficulty using dynamic assessment of algebraic learning. *Journal of Learning Disabilities*, 44, 372–380. doi:[10.1177/0022219411407867](https://doi.org/10.1177/0022219411407867)
- Gielen, S., Dochy, F., & Dierick, S. (2003). Evaluating the consequential validity of new modes of assessment: The influence of assessment on learning, including pre-, post-, and true assessment effects. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new*

- modes of assessment: In search of qualities and standards* (pp. 37–54). Dordrecht: Kluwer Academic.
- Gillam, S. L., Fargo, J., Foley, B., & Olszewski, A. (2011). A nonverbal phoneme deletion task administered in a dynamic assessment format. *Journal of Communication Disorders*, 44, 236–245. doi:10.1016/j.jcomdis.2010.11.003
- Gummersall, D. M., & Strong, C. J. (1999). Assessment of complex sentence production in a narrative context. *Language Speech and Hearing Services in Schools*, 30, 152–164.
- Hamers, J. H. M., Hessels, M. G. P., & Pennings, A. H. (1996). Learning potential in ethnic minority children. *European Journal of Psychological Assessment*, 12, 183–192.
- Haywood, H. C. (2012). Dynamic assessment: A history of fundamental ideas. *Journal of Cognitive Education and Psychology*, 11, 217–229. doi:10.1891/2F1945-8959.11.3.217
- Hessels, M. G. P. (1993). *Leertest voor ethische minderheden: Theoretische en empirische verantwoording* [Learning potential test for ethnic minorities: Theoretical and empirical foundation]. Rotterdam: RISBO.
- Hessels, M. G. P. (2000). The learning potential test for ethnic minorities (LEM): A tool for standardized assessment of children in kindergarten and the first years of primary school. In C. S. Lidz & J. Elliott (Eds.), *Dynamic assessment: Prevailing models and applications* (pp. 109–131). New York, NY: Elsevier.
- Hessels-Schlatter, C., & Hessels, M. G. P. (2009). Clarifying some issues in dynamic assessment: Comments on Karpov and Tzuriel. *Journal of Cognitive Education and Psychology*, 8, 246–251.
- Jeltova, I., Birney, D., Fredine, N., Jarvin, L., Sternberg, R. J., & Grigorenko, E. L. (2011). Making instruction and assessment responsive to diverse students' progress: Group-administered dynamic assessment in teaching mathematics. *Journal of Learning Disabilities*, 44, 381–395. doi:10.1177/0022219411407868
- Jitendra, A. K., & Kameenui, E. J. (1996). Experts' and novices error patterns in solving part-whole mathematical word problems. *Journal of Educational Research*, 90, 42–51.
- Kanevsky, L., & Geake, J. (2004). Inside the zone of proximal development: Validating a multifactor model of learning potential with gifted students and their peers. *Journal for the Education of the Gifted*, 28, 182–217.
- Katz, N., Golstand, S., Bar-Ilan, R. T., & Parush, S. (2007). The dynamic occupational therapy cognitive assessment for children (DOTCA-ch): A new instrument for assessing learning potential. *American Journal of Occupational Therapy*, 61, 41–52.
- Kirkwood, M. W., Weiler, M. D., Bernstein, J. H., Forbes, P. W., & Waber, D. P. (2001). Sources of poor performance on the rey-osterrieth complex figure test among children with learning difficulties: A dynamic assessment approach. *Clinical Neuropsychologist*, 15, 345–356.
- Larsen, J. A., & Nippold, M. A. (2007). Morphological analysis in school-age children: Dynamic assessment of a word learning strategy. *Language, Speech, and Hearing Services in Schools*, 38, 201–212.
- Lauchlan, F., & Elliott, J. (2001). The psychological assessment of learning potential. *British Journal of Educational Psychology*, 71, 647–665.
- Lidz, C. S. (1995). Dynamic assessment and the legacy of L.S. Vygotsky. *School Psychology International*, 16, 143–153.
- Lidz, C. S., & 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, 74–96.
- Meijer, J. (2001). Learning potential and anxious tendency: Test anxiety as a bias factor in educational testing. *Anxiety, Stress & Coping*, 14, 337–362.
- Meijer, J., & Elshout, J. J. (2001). The predictive and discriminant validity of the zone of proximal development. *British Journal of Educational Psychology*, 71, 93–113.
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18, 5–11.
- Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice*, 14, 5–8.
- Minnaert, A., & Vermunt, J. D. (2006). 25 jaar Onderwijspsychologie in Nederland en Vlaanderen in de periode 1980 tot 2005: Trends, pendels en grensverleggers [25 years of Educational psychology in The Netherlands and Flanders during 1980 till 2005]. *Pedagogische Studiën*, 83, 260–277.

- Nirmalakhandan, N. (2009). Use of computerized dynamic assessment to improve student achievement: Case study. *Journal of Professional Issues in Engineering Education and Practice*, 135, 75–80.
- Peltenburg, M., van den Heuvel-Panhuizen, M., & Doig, B. (2009). Mathematical power of special-needs pupils: An ICT-based dynamic assessment format to reveal weak pupils' learning potential. *British Journal of Educational Technology*, 40, 273–284.
- Peltenburg, M., van den Heuvel-Panhuizen, M., & Robitzsch, A. (2010). ICT-based dynamic assessment to reveal special education students' potential in mathematics. *Research Papers in Education*, 25, 319–334. doi:[10.1080/02671522.2010.498148](https://doi.org/10.1080/02671522.2010.498148)
- Perry, N., Turner, J. C., & Meyer, D. K. (2006). Classrooms as contexts for motivated learning. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (2nd ed., pp. 327–348). Mahwah, NJ: Lawrence Erlbaum.
- Prince, A., Minnaert, A., & Opdenakker, M.-C. (2014) *Motivation in educational interventions: A systematic review exploring the concept in relation to intervention outcomes*. Manuscript submitted for publication.
- Resing, W. C. M. (1990). *Intelligentie en leerpotentieel. Een onderzoek naar het leerpotentieel van jonge leerlingen uit het basis- en speciaal onderwijs* [Intelligence and learning potential. A study on the learning potential of young children from primary schools and special education]. Lisse: Swets & Zeitlinger.
- Resing, W. C. M. (2006). Using children's 'ability to learn' in diagnosis and assessment. *Educational & Child Psychology*, 23, 6–11.
- Resing, W. C. M., Tunteler, E., de Jong, F. M., & Bosma, T. (2009). Dynamic testing in indigenous and ethnic minority children. *Learning and Individual Differences*, 19, 445–450. doi:[10.1016/j.lindif.2009.03.006](https://doi.org/10.1016/j.lindif.2009.03.006)
- Ruijsenaars, A. J. J. M. (2001). *Leerproblemen en leerstoornissen. Remedial teaching en behandelingshulp. Hulpschema's voor opleiding en praktijk* [Learning difficulties and learning disabilities. Remedial teaching and treatment. Frameworks for education and practice]. Rotterdam: Lemniscaat.
- Ryba, K. (1998). Dynamic assessment and programme planning for students with intellectual disabilities. *New Zealand Journal of Psychology*, 27, 3–10.
- Schlatter, C., & Büchel, F. P. (2000). Detecting reasoning abilities of persons with moderate mental retardation: The analogical reasoning learning test (ARLT). In C. S. Lidz & J. Elliott (Eds.), *Dynamic assessment: Prevailing models and applications* (pp. 155–186). New York, NY: Elsevier.
- Shamir, A., & Tzuriel, D. (2004). Children's mediational teaching style as a function of intervention for cross-age peer-mediation. *School Psychology International*, 25, 59–78.
- Shapiro, E. S. (1987). *Behavioral assessment in school psychology*. Hillsdale, NJ: Lawrence Erlbaum.
- Shapiro, E. S., & Kratochwill, T. R. (2000). *Behavioral assessment in schools. Theory, research, and clinical foundations*. New York, NY: Guilford Press.
- Sternberg, R. J., Grigorenko, E. L., Ngorosho, D., Tantufuye, E., Mbise, A., Nokes, C., Jukes, M., & Bundy, D. A. (2002). Assessing intellectual potential in rural Tanzanian school children. *Intelligence*, 30, 141–162.
- Swanson, H. L. (1995a). Effects of dynamic testing on the classification of learning-disabilities – The predictive and discriminant validity of the Swanson cognitive processing test (S-cpt). *Journal of Psychoeducational Assessment*, 13, 204–229.
- Swanson, H. L. (1995b). Using the cognitive processing test to assess ability: Development of a dynamic assessment measure. *School Psychology Review*, 24, 672–693.
- Swanson, H. L. (2011). Dynamic testing, working memory, and reading comprehension growth in children with reading disabilities. *Journal of Learning Disabilities*, 44, 358–371. doi:[10.1177/0022219411407866](https://doi.org/10.1177/0022219411407866)
- Swanson, H. L., & Howard, C. B. (2005). Children with reading disabilities: Does dynamic assessment help in the classification? *Learning Disability Quarterly*, 28, 17–34.
- Teo, P., & Roodenburg, J. (2013). Extensions of cognitive ability assessment with dynamic testing using self-regulated learning. *Journal of Cognitive Education and Psychology*, 12, 230–250. doi:[10.1891/2F1945-8959.12.2.230](https://doi.org/10.1891/2F1945-8959.12.2.230)

- Tiekstra, M., Hessels, M. G. P., & Minnaert, A. E. M. G. (2009). Learning capacity in adolescents with mild intellectual disabilities. *Psychological Reports, 105*, 804–814. doi:10.2466/PRO.105.3.804-814
- Tzuriel, D. (2000). The seria-think instrument: Development of a dynamic test for young children. *School Psychology International, 21*, 177–194.
- Tzuriel, D., & Kaufman, R. (1999). Mediated learning and cognitive modifiability: Dynamic assessment of young Ethiopian immigrant children to Israel. *Journal of Cross-Cultural Psychology, 30*, 359–380.
- UNESCO. (1994). *The Salamanca statement and framework for action on special needs education*. Paris: Author.
- Vermunt, J. D. (1996). Metacognitive, cognitive and affective aspects of learning styles and strategies: A phenomenographic analysis. *Higher Education, 31*, 25–50.
- Viljoen, J. L., Odgers, C., Grisso, T., & Tillbrook, C. (2007). Teaching adolescents and adults about adjudicative proceedings: A comparison of pre- and post-teaching scores on the MacCAT-CA. *Law and Human Behavior, 31*, 419–432.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Wigfield, A., Byrnes, J. P., & Eccles, J. S. (2006). Development during early and middle adolescence. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (2nd ed., pp. 87–113). Mahwah, NJ: Lawrence Erlbaum.
- Wilson, B. (1998). Matching student needs to instruction. In S. A. Vogel & S. Reder (Eds.), *Learning disabilities, literacy and adult education* (pp. 191–212). Baltimore, MD: Brookes.
- Woide, N., Beckmann, J. F., Elliott, J. G., & Guthke, J. (2005). *Effects of feedback: Individual differences in the ability to benefit from feedback*. Paper for the 10th Conference of the International Association on Cognitive Education and Psychology, University of Durham, England.