

Dynamic Testing

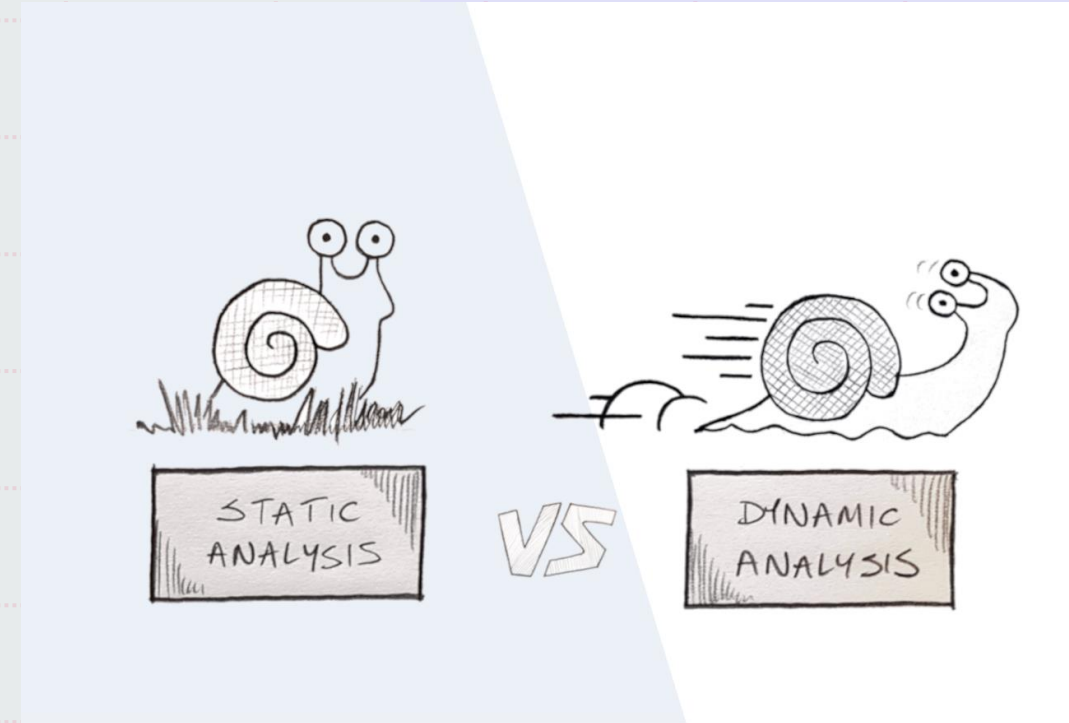
by Elena Grigorenko & Robert Sternberg
Yale University
Psychological Bulletin (1998)

Paper Review By: Christian Hahm

Temple AGI Team

Introduction

- Traditional measures of cognitive ability (e.g., IQ test) measure *realized skills*, not *latent cognitive potential*. Such **static tests** are problematic when it comes to achieving their goal of measuring raw intelligence, because:
 - 1.) *cognitive potential*, not specific skill, is often the subject of interest when it comes to quantifying “intelligence”
 - 2.) complex skills do require high cognitive potential, but the acquisition (or lack of acquisition) of these skills is influenced by external factors: provided education, upbringing, etc.
- **Dynamic testing**, as opposed to static tests, is proposed as a way to measure learning potential. They differ by:
 1. Static testing measures the products of learning, dynamic testing instead measure learning while it occurs.
 2. Static testing presents a set of tasks to be solved. During dynamic testing, the test proctor provides verbal feedback to the test taker during each task. (the test itself is a pairing of pure evaluative testing + instruction)
 3. The role of test proctor changes from “neutral scientific evaluator” to “interactive teacher”



Dynamic Testing

- Based on both learning and the products of learning
- Consists of multiple tasks to solve
- Assumes the test taker starts with almost no knowledge of each task, but all necessary knowledge can be learned by using the teacher
- The goal: whether and how will the subject change their behavior if provided the opportunity?
- The “teacher’s feedback” can range from simple & standardized to highly targeted for the test-taker.





The Need for Dynamic Testing and the Attempts to Meet this Need

Rationale

- Measure *learning as it occurs* to evaluate learning ability
- Reduce unfair outcomes in standardized testing, by providing all students equal opportunities to succeed on the test itself
- Predict specific fields for which a person has aptitude (e.g., they more efficiently learn biology than chemistry)
- *Example:* a static test on calculus. What if a student could easily master calculus in a few days, but simply was never taught?
- Static testing measures the amount of currently attained knowledge about calculus
- Dynamic testing measures the potential for mastering calculus



History: Zone of Proximal Development

- Lev Vygotsky's school of dynamic testing - Zone of Proximal Development (ZPD)
- Study children, they are the simplest and clearest examples of learners, since they are constantly learning as they mature
- Test what a child can do independently (**actual** level of development), and compare that to what they can do with adult help (**possible** level of development); the distance between these (**possible** MINUS **actual**) is the ZPD
- The ZPD gap closes as the student gains independent proficiency, such as when taught by a teacher about how to think about the subject generically.



Leading Modern Approaches to Dynamic Testing

(Survey)

The Approaches in Action

4 major clusters of Dynamic Learning paradigms:

- Meta-cognitive intervention targeted at teaching generalizable concepts and principles
- Learning within the test
- Restructuring the test situation
- Training a single cognitive function

4-point assessment of each approach:

- *Comparative Informativeness*
- *Power of Prediction*
- *Degree of Efficiency*
- *Robustness of Results*

Properties

Table 1
Dynamic-Testing Approaches

Approach	Method	Target population	Format	Context of testing (nature of the task)	Outcome (goal)	Focus (orientation)	Predictive power
Theory of structural cognitive modifiability (Feuerstein & colleagues)	Learning Potential Testing Device	All individuals who can use modification	Test—mediate—test	Artificial (outside the context of school program)	Structural cognitive changes	Child-driven	Not well established
Learning potential testing (Budoff & colleagues)	Test-centered coaching	Children who have experienced school failure (low-IQ, low-achieving students)	Formal pretest—standardized training/coaching—formal posttest	Artificial (outside the context of school program); abstract reasoning problems (mostly nonverbal)	Improved test performance	Task-driven	Fairly high
Testing through learning and transfer (graduated-prompts approach; Campione & Brown)	Hinting procedure	Scholastically weak students	Pretest (level-of-performance information)—initial mediated learning—static maintenance and transfer testing—mediated maintenance and transfer	With the exception of usage of traditional tests, testing is situated within specific domains	A measure of zone of proximal development	Task-driven	Not established
Lerntest approach (learning potential testing)	German Learning Potential Tests (Guthke & colleagues)	Normal, mentally retarded children and adults with brain disorders	(a) pretest—training—posttest (long-term) and (b) train-within-test paradigm (short-term)	Psychometrically oriented approach where testing is situated within specific domains	Records of learning gain	Task-oriented	Fairly high for individuals with IQ below average
	Dutch Learning Potential Test for minority groups (Heesels & Hamers)	Children of ethnic minorities	Train-within-test paradigm	Psychometrically oriented approach where testing is situated within specific domains	Extent to which children benefit from help	Task-oriented	The test appears to be a moderate predictor of school achievement, but not a better predictor than a static intelligence test
Testing-the-limits approach	Teach-to-the limit approach (Carlson & Wiedl)	Normal, mentally retarded, and learning-disabled children	Multiple conditions (varying amount of verbalization and feedback)	Testing is situated within specific domains	Improved test performance	Task-driven	No better than IQ in terms of prediction of school achievement, but is predictive of adaptive teaching styles
Information-processing framework	Swanson's Cognitive Processing Test (S-CPT)	Learning-disabled children	Test—teach—test	Artificial (working memory tasks)	Indicator of processing potential	Task-driven	Appears to be fairly high for learning-disabled children, especially the group of slow learners

Feuerstein's Approach

- The "Mediated Learning Experience" to measure *cognitive modifiability*
 - An adult mediates between the child and the task, adjusting task order / complexity or stimulating the child's attention
- The "Learning Potential Assessment Device" (LPAD)
 - Meant to assess potential in people of any age
 - First, guided exposure to problems; then, independent exposure
 - The examiner uses an array of 15 tasks to trigger certain cognitive functions; then they create a cognitive map to identify the areas of deficit and how modifiable each area is
 - Whereas static tests only identify deficit of cognitive functions, the dynamic tests lets the examiner determine "pliability" of those functions and focus on those which are especially pliable to overcome the deficits
- This endeavor inspired many other works on dynamic testing, but has strong criticisms:
 - The 15 tasks are static tests, so the "dynamicness" relies most heavily on the mediator
 - There is no scientific consensus for the exact cognitive functions measured in the test
 - Most studies could not find evidence of LPAD's efficacy, and the ones that did are questionable



Budoff's Approach to Measuring Learning Potential

Meant for disadvantaged children (underachiever, learning-disabled); Budoff thought with specialized instruction, some could improve

Budoff et al. modified standardized tests to be dynamic, and there is a specific set of instructions for the teacher.

Pretest stage: administer static test as baseline

Training stage: direct the student's attention, explain crucial attributes of the task and the test, and guide the student through the correct actions to find the solution

Testing stage: 1 day later, then 1 month later, re-administer static test

Learning potential is something like *post test score MINUS pre test score*

They classified the children into 3 groups 1.) high scorers , 2.) gainers, 3.) non-gainers

The biggest success here is that *static testing* places groups 2 and 3 as identical (low scoring), whereas *Budoff's approach* differentiates groups 2 and 3 to identify students which demonstrate learning potential but do not perform well for some reason in school

Criticisms:

Too population-specific (disadvantaged children)

Not enough evidence that these 3 groups can predict future success better than traditional IQ measures



Testing by Learning and Transfer (the Graduated-Prompts approach)

- This approach establishes a supportive framework that gradually helps the test-taker until they can solve the test problem. No direct teaching is involved, only hints
- The key idea here is that they test transfer, using knowledge from one context in different contexts.
- Procedure; several sessions with hinted and non-hinted stages. In the hinted stage, more and more specific hints are gradually given as the test-taker struggles or fails, until they succeed.
 - Initial static pre-test
 - Initial learning, with hints
 - Static post-test (tests both maintenance and transfer)
 - Post-test with hints
- Instead of measuring difference in performance, they measure learning potential as the inverse of the minimal number of hints required by the test-taker (e.g., $1/2$ hints = 0.5, $1/3=0.333$, etc.), the hints are summed across the stages . This is a measure of how much help the test-taker needed, a measure of learning potential
- Did predict future academic achievement well.
Criticism: hints are hard to quantify, some are more helpful than others; how to know whether the difference between hints is additive, hint 3-2 is the same as hint 2-1?



The European Contribution



- Guthke's approach: the *Lerntest*; pretest – training – posttest
 - Pretest to gain familiarity with the problems
 - During training:
 - repeat easier-class problems when first failing harder-class problems
 - If failing the problem again, receive hints or prompts to guide to the solution
 - The learning potential is the post-test score
- Carlson and Wiedl – *Testing-the-limits* approach
 - They attribute much of poor performance to the test-taker's inability to understand the problem and to contributing individual personality traits (test anxiety)
 - The goal is to manipulate the questions so as to compensate for the test-taker's educational deficit to get the best possible performance
 - During the test, the examiner prods the test taker with structured verbal questions ("what are you thinking about as you read this problem? Why do you think X answer is correct compared to the others?") and feedback when the test-taker makes mistakes

Swanson's Cognitive Processing Test

- Main assumption is that working memory is critical in learning
- Swanson defined:
 - Working memory as simultaneously holding old and new data to manipulate or transform it
 - Long-term memory as a system of interconnected units representing semantic and episodic content
- 11 subtests, to measure 7 various *processing potential* (similar to Feuerstein's *cognitive modifiability*) scores (initial, gain, probe, maintenance, processing difference, processing stability, strategy efficiency)





A 3-Prong Approach to Dynamic Testing

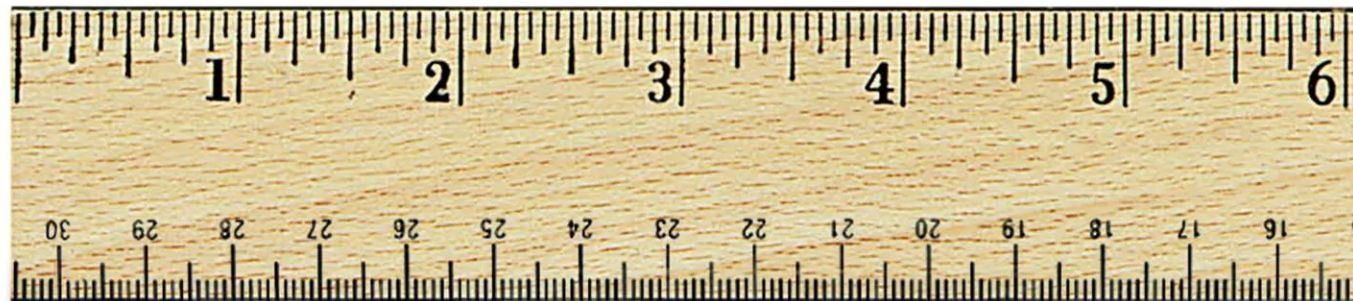
1.) General Aim: evaluate, modify, or both?

- 3 different goals of dynamic testing:
 - Provide a better metric of test-takers' intelligence by accounting for educational disparities
 - Measure the acquisition of new skills / cognitive modifiability of the test-taker
 - To improve mental efficiency (change the test-taker's level of ability)
- All goals are valid; which goal to use dynamic testing for will determine the nature of the dynamic test (do you want to improve your student's performance overall or just enough to measure it?)



2.) Measurement of Change

- Biggest problem is that there is no consensus on the psychometrics to use; is post-test minus pre-test really a good measure of change?
- It is assumed the student's got better because of the teacher's instruction, but it was found in large parts of improvement could be accounted for because the students were given a second chance at the class of task.
- There is critique to pure quantitative approach and some calls for binary qualitative approach, to just determine whether the test-taker changed favorably or not.



3.) Ecological Validity of Dynamic Testing

- Even though these tests claim to measure change, they are still only measuring the products of cognition (post-test, pre-test) not the process itself. Therefore, the tests are intrinsically only valid in terms of product-based criteria, like school tests
- Despite being interesting alternatives to static testing, they do not show greater predictive power (such as school achievement) than the static tests. Perhaps this is because school achievement itself is measured in product-based criteria as well
- There is a mismatch between the process being measured (the test-taker's learning process) and the test process itself, which frames itself as passively being interacted with by the test-taker, rather than constantly actively probing and measuring the learning process



End