## PART ONE. THEORY AND PRINCIPLES

# 1 Dynamic Assessment: Introduction and Review

### DEFINITIONS

Dynamic assessment (DA) is no longer a new approach to psychological and educational assessment; in fact, some of its current applications have been around for more than a half century (see, e.g., Feuerstein, Richelle, & Jeannet, 1953; Guthke & Wingenfeld, 1992). Despite such a relatively long history, it is still not widely practiced around the world (Elliott, 1993; Lidz, 1991, 1992). In April 2005, 588 literature citations relating to dynamic assessment were listed at the Web site www.dynamicassessment.com. The majority of those are of recent date, suggesting a rapid growth of interest in this topic in the last 10 to 15 years. A much broader search engine (www.google.com) produced 17,800,000 hits for this term; to be sure, the overwhelming majority of these did not relate to "dynamic assessment of learning potential."

At the dynamic assessment Web site, DA is defined as "an interactive approach to conducting assessments within the domains of psychology, speech/language, or education that focuses on the ability of the learner to respond to intervention." Others have defined it variously, but the constant aspect of the definition is active intervention by examiners and assessment of examinees' response to intervention. Haywood (1992b) suggested that dynamic assessment is a subset of the more generic concept of interactive assessment. He further suggested that "It might be useful to characterize as interactive any approach to psychological or psychoeducational assessment in which the examiner is inserted into an active relationship with a subject and does more than give instructions, pose questions, and record responses. 'Dynamic' should probably be reserved for those approaches in which the interaction is richer, in which there is actual teaching (not of answers but of cognitive tools), within the interaction and in which there is conscious, purposeful, and deliberate effort to produce change in the subject" (Haywood, 1992b, p. 233). Haywood and Tzuriel (2002) defined

1

dynamic assessment as "a subset of interactive assessment that includes deliberate and planned mediational teaching and the assessment of the effects of that teaching on subsequent performance" (p. 40). In current use, the two terms appear to be used interchangeably, together with such others as "dynamic testing" (e.g., Sternberg & Grigorenko, 2002; Wiedl, 2003) and "learning potential assessment" (e.g., Budoff, 1987; Hamers, Sijtsma, & Ruijssenaars, 1993). For further definition, see Carlson and Wiedl (1992a, 1992b), Feuerstein, Rand, and Hoffman (1979), Guthke and Wingenfeld (1992), Haywood and Tzuriel (1992), Lidz (1987), and Tzuriel (e.g., 2001). All of these approaches are in some sense "mediational," but there are other approaches to assessment that include intervention and response to intervention but that are not mediational. These would fit within the broad definition of DA.

## Applicability of Dynamic Assessment

Although a few authors have suggested that dynamic assessment of learning potential should replace standardized, normative intelligence testing, our position does not come close to that. In fact, we insist that DA is not for everybody on all occasions but instead constitutes a valuable part of the assessment repertoire when used in conjunction with other forms of assessment, including standardized testing, social and developmental history taking, observation of performance in learning situations, and data gathered from clinical interview, parents, teachers, and others. The DA part of the repertoire is needed because it can add information about both present and potential performance that is not readily (or even at all) obtainable from other sources. Most dynamic assessment experts (e.g., Feuerstein, Haywood, Rand, Hoffman, & Jensen, 1982/1986; Haywood, 1997; Lidz, 1991) have suggested that this method is especially useful when

*scores on standardized, normative tests are low*, and especially when they do not accord with information from other sources;

*learning appears to be restrained* by apparent mental retardation, learning disability, emotional disturbance, personality disorder, or motivational deficit;

*there are language problems*, such as impoverished vocabulary, difference between the maternal language and the language of the school (workplace), or delays in language development;

*there are marked cultural differences* between those being examined and the majority or dominant culture, as, for example, in recent immigrants; and

*classification is not the only or central issue, but the need to inform programming is important.*

In all of these situations, standardized, normative testing is likely to yield low scores and consequent pessimistic predictions of future learning effectiveness and school achievement. It is not the major role of DA to dispute those predictions; indeed, they are disastrously likely to prove accurate if nothing is done to overcome various obstacles to learning and performance. The role of DA is rather to identify obstacles to more effective learning and performance, to find ways to overcome those obstacles, and to assess the effects of removal of obstacles on subsequent learning and performance effectiveness. By extension of that role, a goal of DA is to suggest what can be done to defeat the pessimistic predictions that are often made on the basis of results of standardized, normative tests, including estimating the kinds and amount of intervention that will be necessary to produce significant improvement and the probable effects of such intervention. At the present stage of development of DA, those estimates are only rarely reducible to numbers. In fact, many adherents to DA resist precise quantification of estimates of learning potential or of ability to derive benefit from teaching because of fear that such quantification could lead to the use of DA data for the same purposes for which normative, standardized testing is generally used: classifying people, identifying those who are not expected to do well in school and other educational settings, and rank-ordering people with respect to presumed intellectual ability. One imagines with horror the development and use of a "modifiability index" or "learning potential quotient"!

## Comparison of Dynamic and Normative/Standardized Assessment

A recurring theme in this volume is that the psychoeducational assessment process relies on data from diverse sources, of which DA is one. Because of that emphasis, it is useful to ask what it is that standardized tests do not do, or do not do well, and how DA can fill the gap left by these tests – or, indeed, to correct some of the errors of assessment that psychologists and others make when they rely exclusively on data from standardized tests.

Dynamic assessment has certain limitations, as well as some yet-unsolved problems, that make it important that the method be used appropriately, for specific purposes. First, because all approaches to dynamic assessment involve some effort to change examinees' performance, the data should not be used for classification and cannot be referred to normative tables for interpretation. Second, much of the interpretation of DA data depends on the skill and experience of the examiner. Third, the reliability of inferences regarding deficiencies in cognitive functioning has not been

well established; that is, different examiners may reach different conclusions that reflect their own training and experience. In fact, the particular model that we present in this volume is not a deficit model; rather, it is one in which DA is used principally to identify areas of strength and potential strength, to discover what performance might be possible given optimal learning conditions and appropriate intervention, and to specify what those optimal conditions might be.

Standardized tests of intelligence are excellent instruments for the general purpose of classification, which is a useful activity when one is planning the allocation of limited resources or attempting to place individuals in groups where they can be served effectively. One such use is to identify gifted and talented children and youth so they can be educated in classes that require more investment of intellectual resources than do average public school classes. Another is to identify persons at the other end of the IQ distribution, that is, those who are likely to be mentally retarded and to require or benefit from special educational services.

By comparing each individual's score on standardized intelligence tests with the average score of persons of similar age and social background, that is, with the norms of the tests, one essentially rank-orders the tested person with respect to persons in the normative samples. One is then able to make such statements as, "This person's intellectual functioning is below that of 95% of children of his age." Even when tempered with a probability statement such as, "There is a 95% chance that his intellectual development lies between the 5th and 15th percentiles," it is still a rather confident statement that says there are severe limits on what can be expected by way of school learning. What is more, such an exercise tells us about an individual's performance compared with that of groups of others but nothing at all about how that performance could be enhanced.

The correlation between standardized intelligence test scores (IQ) and subsequent school achievement scores is between +.55 and +.80 – usually considered a strong correlation. Taking a value that is often cited, +.70, and squaring that coefficient, we find that IQ and subsequent school achievement share only 49% common variance, leaving roughly half of the variance in school achievement to be associated with other variables. For present purposes, what that means is that there is substantial error in using IQ to predict subsequent school achievement of individuals, and even of large groups; therefore, the usefulness of IQ even for classification is limited.

Returning to our illustration of "gifted and talented" and "mentally retarded" persons, a common observation is that the predictive errors are made in opposite directions. That is to say, relatively few people would

be included in the "gifted and talented" group, so identified, who did not belong there, but there might well be a much larger number who would do well in special "gifted" classes but whose test scores do not qualify them for that category. On the other hand, overinclusion is the more common error when constituting groups of persons with mental retardation, resulting in the assignment of relatively many individuals to special classes for children with intellectual disability who might do better in regular, mainstreamed classes. In other words, standardized intelligence tests are more likely to make individuals appear to be less intelligent than they are capable of being than to make them appear to be more intelligent. How much difference that relatively constant error makes depends on whether one is working at the top or the bottom of the distribution of intelligence.

There are other important differences, and these are summarized in Table 1.1. It is important to note that our focus in this book is the presentation of our own approaches to DA. Because our DA roots are in mediational theory and practices, there is an inevitable bias in this direction.

Utley, Haywood, and Masters (1992), like Jensen (1980), found no convincing evidence to support the claim that standardized tests are inherently biased against certain subgroups in the population, such as ethnic minorities. Reviewing available literature on psychological assessment of minority children, they concluded

> that psycho-educational assessment instruments on which minority and majority groups score differently are valid according to a variety of criteria that are relevant to the tests' theoretical underpinnings. It might be said that such instruments actually have several kinds of validity, one of which is validity with respect to the use to which the tests [are] put. Tests that yield an intelligence quotient might possess strong validity in terms of being able to predict aspects of performance and achievement that can be linked to the concept of intelligence, but at the same time they might have poor validity in predicting responsiveness to a particular educational regimen that adapts teaching to meet certain needs. Put differently, a test that is used to assess how well or how rapidly a child learns may not predict how that child might best be taught. For such a purpose the best assessment might be one that targets how a child learns so that instruction may be tailored either to the child's manner of learning or toward changing how the child learns. In short, there is validity-for-a-given-purpose, and an instrument that is valid for one purpose (e.g., predicting correlates of intelligence) may not be valid for another (predicting the best sort of educational experience). (1992, p. 463)

**Table 1.1.** Comparison of "normative" and "dynamic" assessment approaches[a]

| Comparison criterion what is compared | Normative assessment self with others | Dynamic assessment self with self |
|---|---|---|
| Major question | How much has this person already learned? What can he/she do or not do? How does this person's current level of performance compare with others of similar demographics? | How does this person learn in new situations? How, and how much, can learning and performance be improved? What are the primary obstacles to a more optimal level of competence? |
| Outcome | IQ as global estimate of ability *reflecting rank order in a reference (normative) group* Current level of independent functioning (ZOA) | Learning potential: What is possible with reduced obstacles to learning? How can such obstacles be reduced? How does the individual function with the support of a more experienced interventionist? (ZPD) |
| Examining process | Standardized; same for everybody Focus on products of past experience | Individualized; responsive to person's learning obstacles Focus on processes involved in intentional acquisition of new information or skills |
| Interpretation of results | Identification of limits on learning and performance; identification of differences across domains of ability Documentation of need for further assessment and possible intervention | Identification of obstacles to learning and performance; estimate of investment required to overcome them Hypotheses regarding what works to overcome obstacles to learning |
| Role of examiner | Poses problems, records responses; affectively neutral | Poses problems, identifies obstacles, teaches metacognitive strategies when necessary, promotes change; affectively involved |

ZOA = zone of actual development; ZPD = zone of proximal development.
[a] Adapted from Feuerstein, Haywood, Rand, Hoffman, and Jensen (1982/1986), and from Haywood and Bransford (1984).

Utley, Haywood, and Masters (1992) further concluded that

> (a) standardized intelligence tests are not reasonably called upon to
> do jobs that we now see as important; (b) ethnic minorities may be
> especially subject to erroneous decisions and placements based upon
> standardized intelligence tests, not because of test bias or poor pre-
> dictability but because ethnically and culturally different persons might
> often have need of different educational approaches that are not iden-
> tified by standardized normative tests; (c) these are legitimate public
> policy issues; and (d) dynamic assessment has the potential to be an
> important adjunct to standardized intelligence tests, especially for use
> with ethnic minorities and other persons who are socially different,
> such as handicapped persons, culturally different persons, and persons
> whose primary language is other than that of their dominant culture.
> (1992, pp. 463–464)

These observations are in accord with our own position, the focus of which
is on *how test data are interpreted and used* for making important decisions
about people's lives. Our objection to exclusive reliance on intelligence
tests for data to inform such decisions is primarily that intelligence test
data are remarkably subject to misuse, whereas DA data can supply what
is missed in the testing of intelligence.

## CONCEPTS, ASSUMPTIONS, AND THEORETICAL BASIS
## OF DYNAMIC ASSESSMENT

Some fundamental concepts and assumptions appear to underlie virtu-
ally all approaches to dynamic/interactive assessment. They include the
following:

> 1. Some abilities that are important for learning (in particular) are not
> assessed by normative, standardized intelligence tests.

> 2. Observing new learning is more useful than cataloguing (presumed)
> products of old learning. History is necessary but not sufficient.

> 3. Teaching within the test provides a useful way of assessing potential
> as opposed to performance.

> 4. All people typically function at less than their intellectual capacity.

> 5. Many conditions that do not reflect intellectual potential can and do
> interfere with expression of one's intelligence.

The notion that some important abilities are not typically assessed by nor-
mative, standardized intelligence tests is not worth much unless one can
identify ways to assess those fugitive abilities. One prominent way is to

look for conditions that may be limiting a person's access to his or her intel-
ligence, minimize or remove those limiting conditions, and then assess
abilities again. This strategy is exactly the one that led Vygotsky to his
now-famous concept of the "zone of proximal development":

> Most of the psychological investigations concerned with school learn-
> ing measured the level of mental development of the child by making
> him solve certain standardized problems. The problems he was able
> to solve by himself were supposed to indicate the level of his mental
> development at the particular time. But in this way, only the completed
> part of the child's development can be measured, which is far from
> the whole story. We tried a different approach. Having found that the
> mental age of two children was, let us say, eight, we gave each of them
> harder problems than he could manage on his own and provided some
> slight assistance: the first step in a solution, a leading question, or some
> other form of help. We discovered that one child could, in cooperation,
> solve problems designed for twelve-year-olds, while the other could
> not go beyond problems intended for nine-year-olds. The discrepancy
> between a child's actual mental age and the level he reaches in solving
> problems with assistance indicates the zone of his proximal develop-
> ment; in our example, this zone is four for the first child and one for the
> second. Can we truly say that their mental development is the same?
> Experience has shown that the child with the larger *zone of proximal
> development (ZPD)* will do much better in school. This measure gives a
> more helpful clue than mental age does to the dynamics of intellectual
> progress. (Vygotsky, 1986/1934, pp. 186–187)

Although there have been some improvements recently in intelligence
testing (e.g., Das & Naglieri, 1997; Woodcock, 2002), it remains true that
much of standardized intelligence testing relies on assessment of the prod-
ucts of presumed past learning opportunities. Vocabulary tests, for exam-
ple, are common, and these by their very nature reflect past learning. Com-
prehension of social situations, humor, and absurdity shows up often in
such tests and similarly has to be based on prior learning, as does knowl-
edge of mathematics and skill at calculating. Comparison of any individ-
ual's score on such tests with the average score of similar persons in a
normative sample requires the logical assumption that all persons of a
given age, gender, race, and social circumstance (e.g., urban vs. rural res-
idence) have had the same opportunities to learn – an assumption that is
patently untenable. Although old learning is highly correlated with success
in new learning (the venerable "principle of postremity" in psychology: the
most likely response is the most recent response, or the best predictor of
future behavior is past behavior), the correlation is far from perfect and

often becomes a self-fulfilling prophecy. An obvious example is the deaf child who comes to school without having had the benefits of training in specialized communication. That child will score poorly on normative tests because he or she will have learned less than have age peers, but the score will not reflect the potential of the child to learn *given appropriate communication methods*. In such a case, attempts within the test to overcome experiential deficits will yield better estimates of the child's ability to learn, given appropriate teaching. Teaching within the test should bear greater resemblance to the criterion situation, in this case classroom learning in a person-appropriate class. If, on the other hand, such a child is given normative tests, scores low, and is placed in learning situations with low expectations and without appropriate communication methods, the prophecy of the normative score will be fulfilled because the assessment and criterion situations are similar.

All proponents of dynamic assessment appear to be more interested in determining potential performance than in assessing typical performance. They recognize that all people typically function at levels far below their capacity, at least in terms of their neural capacity. Assessment of typical performance is invaluable for prediction of future performance. If one wishes, however, to assess what is possible or what would be possible under more optimal conditions – in other words, how to defeat pessimistic predictions derived from assessment of typical performance – then a testing strategy that involves intervention and the seeking of potential is essential.

From the beginning of psychological science, psychologists have been careful to distinguish between "intellective" and "non-intellective" variables. Early psychologists, for example, divided "consciousness" into the three dimensions of *cognition* (knowledge, thinking), *conation* (feeling, emotionality, perhaps motivation), and *volition* (will) (Boring, 1950). Whereas such a division makes for good science in the search for pure effects, uncontaminated by "irrelevant" variables, it does not make for good clinical assessment, especially when assessment of intelligence is based heavily on performance on tests that require prior learning. We know, for example, that intelligence test scores can be affected by motivational variables (Zigler, Abelson, & Seitz, 1973; Zigler & Butterfield, 1968), racial, gender, and linguistic match between examiner and examinee, language competence, previous testing experience, social class, personality of examiner and examinee, and a host of other non-intellective variables (see, e.g., Tzuriel & Samuels, 2000; Tzuriel & Schanck, 1994). Almost all DA advocates, then, try to identify and compensate for the effects of such non-intellective variables and to take them into account when interpreting the data from DA. Some typical sources of poor performance that can be easily overcome include misunderstanding of instructions or expectations (Carlson & Wiedl, 1992a,

1992b), unfamiliarity with materials and content of tests, timidity, and history of failure on tests (Johnson, Haywood, & Hays, 1992).

## AN APPROACH TO DYNAMIC ASSESSMENT

Anyone's specific approach to assessment of individual differences in human abilities should derive from and depend on one's systematic view of the nature of human ability itself. We do represent a particular view of that subject, discussed in detail in Chapter 2. We present a synopsis of the applied aspects of that approach here to make it easy to compare it with other approaches within the broad field of DA.

Our approach to dynamic assessment is actually an approach to psychological and psychoeducational assessment in general; that is, we do not separate DA as a complete alternative to more traditional assessment methods. This approach includes the use of DA for the purpose of finding answers to specific questions, as a specific tactic within an assessment strategy that includes more traditional methods, such as standardized testing.

In general, we find that the social–developmental history is the single most important source of diagnostic information, and it often contains clues to fruitful intervention strategies as well. Careful history taking, supplemented by records review, interview, and direct observation in learning and performance situations, is the primary source of the questions to be addressed in the more formal aspects of assessment. The nature of those questions must determine which tactics to employ and in what sequence.

The first major aspect of our approach to DA is our answer to the question, "Why do dynamic assessment?" We distinguish the principal goals of DA from those of static, normative assessment along two axes. The first is to consider what one seeks to find out from administering ability tests, that is, what question(s) one asks of the assessment process. "Intelligence" tests, although initially designed simply to sort out children who might or might not be expected to succeed in "regular" classes at school (Haywood & Paour, 1992), have nevertheless come to be seen as instruments for making inferences about a latent variable – intelligence – that is not observable and not measurable by any direct means. Doing so is important within the context of development and elaboration of theories of human development and functioning. That is not a goal of DA, in which one seeks instead to make inferences about barriers to the expression of intelligence and ways to improve functioning, especially in the sphere of learning, both academic and social. A second goal of standard intelligence tests is classification: placing into categories or ranges of intelligence those persons who score at certain IQ levels. This is done on the assumption that persons who achieve similar scores on the tests have enough characteristics