

**Dynamic Assessment of L2 Development:
Bringing the Past into the Future**

**Matthew E. Poehner
James P. Lantolf**
The Pennsylvania State University

CALPER Working Paper No. 1
October 2003

Dynamic Assessment of L2 Development: Bringing the Past into the Future*1

Matthew E. Poehner
James P. Lantolf
The Pennsylvania State University

Introduction

The present paper outlines a theoretical framework for a research program on Dynamic Assessment (henceforth, DA) within in the fields of L2 research, pedagogy and language testing. To achieve this, we will first discuss the theoretical basis of DA in the work of L. S. Vygotsky; next, we will contrast DA with more traditional static approaches to assessment (henceforth, SA) in the general educational and psychological literatures; we will then review the few studies that have been carried out to date on DA and L2 learning and instruction; we will next consider some of the critiques leveled against DA, in particular in its clinical orientation, by those concerned with psychometric principles; finally, we will consider the implications of some recent theoretical and empirical research calling for a closer connection between L2 assessment and instruction in light of our discussion of DA.

The terms DA and SA were formulated by researchers working in the DA paradigm as a way of distinguishing what they see as two radically different models of assessment. Sternberg and Grigorenko (2002) characterize SA as follows:

The examiner presents items, either one at a time or all at once, and each examinee is asked to respond to these items successively, without feedback or intervention of any kind. At some point in time after the administration of the test is over, each examinee typically receives the only feedback he or she will get: a report on a score or set of scores. By that time, the examinee is studying for one or more future tests. (vii)

In contrast, the authors define DA as a procedure whose outcome

takes into account the results of an intervention. In this intervention, the examiner teaches the examinee how to perform better on individual items or on the test as a whole. The final score may be a learning score representing the difference between pretest (before learning) and posttest (after learning) scores, or it may be the score on the posttest considered alone. (ibid.)

* Research for this article was funded in part by a grant from the United States Department of Education (CFDA 84.229, P229A020010-03). However, the contents do not necessarily represent the policy of the Department of Education, and one should not assume endorsement by the Federal Government. It was also partly funded by a Gil Watz Fellowship from the Center for Language Acquisition at The Pennsylvania State University. We would like to thank Steve Thorne and Merrill Swain for their helpful feedback on earlier versions of the manuscript.

From these definitions it is clear that DA focuses "on modifiability and on producing suggestions for interventions that appear successful in facilitating improved learner performance" Lidz (1991: 6), while in SA no attempt is made to change the examinee's performance (Haywood, Brown & Wingenfeld 1990).

Dynamic Assessment and the Zone of Proximal Development

Vygotsky's writings on the Zone of Proximal Development (ZPD) provide the theoretical underpinnings of DA. Central to the ZPD, and the core concept of Vygotsky's theory of mind, is *mediation*. Higher forms of thinking are socially and culturally derived, emerging as a consequence of our interactions with other individuals and with physical and symbolic artifacts (e.g., books, paper and pencil, computers, diagrams, numbers and language, etc.) constructed by others in different places and at different times. In this way, our relationship to the world is not direct but mediated.

Vygotsky's formulation of the ZPD was based on his observation that schooling frequently enhanced the IQ score of some but not all children (van der Veer and Valsiner 1991). Importantly, it was the children who entered school with low scores who most often improved over time, while those with already high scores generally did not show much improvement. Vygotsky reasoned that this differential effect of schooling was a consequence of the fact that children with high IQs had already traversed the distance between their *actual* and *potential* development (relative to what the school curriculum concretely offered) prior to entering school, but that their low IQ classmates still had room for development to occur.

Vygotsky (1998: 201) argued against the general view that *independent problem solving* was the only valid indication of mental functioning, suggesting instead that this revealed only part of a person's mental ability, his or her *actual* developmental level. Indeed, "determining the actual level of development not only does not cover the whole picture of development, but very frequently encompasses only an *insignificant* [italics added] part of it" (Vygotsky 1998: 200). He insisted that responsiveness to assistance is an indispensable feature for understanding cognitive ability because it provides an insight into the person's *future* development. That is, what the individual is able to do one day with assistance, s/he is able to do tomorrow alone. Importantly, potential development varies independently of actual development, meaning that the latter, in and of itself, cannot be used to predict the former. Moreover, the former is not an a priori prediction but is derived from concrete activity mediated by others or by cultural artifacts. The following quote from Vygotsky provides what we see as an early description of DA:

Imagine that we have examined two children and have determined that the mental age of both is seven years. This means that both children solve tasks accessible to seven-years-old. However, when we attempt to push these children further in carrying out the tests, there turns out to be an essential difference between them. With the help of leading questions, examples, and demonstrations, one of them easily solves test items taken from two years above the child's level of [actual] development. The other solves test items that are only a half-year above his or her level of [actual] development. (Vygotsky 1956: 446-447, cited in Wertsch 1985: 68)

For Vygotsky the two children are simultaneously equivalent and not equivalent, as he points out a bit further in the same work:

From the point of view of their independent activity they are equivalent, but from the point of view of their immediate potential development they are sharply different. That which the child turns out to be able to do with the help of an adult points us toward the zone of the child's proximal development. This means that with the help of this method, we can take stock not only of today's completed process of development, not only the cycles that are already concluded and done, not only the processes of maturation that are completed; we can also take stock of processes that are now in the state of coming into being, that are only ripening, or only developing (Vygotsky 1956: 447-448; cited in Wertsch 1985: 68)

It is clear from Vygotsky's characterization of the ZPD that to fully assess an individual's development, it is not enough to determine her or his intrapsychological ability, we must also uncover her or his interpsychological capacity. Said another way, observing a person's history (i.e., actual level of development) presents only part of the picture; the full picture emerges when we take account of his or her future.

Contrasting Dynamic Assessment and Static Assessment

Understanding the Future

Vygotsky's theorizing in the ZPD is predicated upon a radically different understanding of the future from that which informs SA, and in our view this is the fundamental difference between DA and SA. To appreciate this difference we will draw on Valsiner's (2001) work on three ways that developmental psychology has theorized the future. In the first model, essentially exclusionary or atemporal, humans do not develop but rather mature as a consequence of either genetically (e.g., innatism) or environmentally (e.g., behaviorism) specified causal factors. The second and third models are characterized by Valsiner as *past-to-present* and *present-to-future*, respectively. The former recognizes "the role of the past life history of the organism in leading to its present state of functioning" (p. 86). Development in this model is seen as a "sequence of stages" that a person is assumed to pass through on the way to some final stage; moreover, stages cannot be skipped along the way. The future is predicted "*post factum* – when it already has become present" (Valsiner 2001: 86). Valsiner goes on to point out that "the underlying assumption that is axiomatically accepted here is that the dynamic *changes* of the past that have led to the present can also explain any future. History (of the past) is here utilized to eliminate history (of new development) for the future. The future is assumed to be similar to the past" (ibid.), or as Frank Lloyd Wright's famous dictum puts it, "The future is now." Freud's theory of emotional development, Piaget's theory of cognitive development, and, within SLA, Krashen's (1981, 1983) morpheme-order hypothesis and Pienemann's (1998) processability hypothesis are all examples of past-to-present models of development.

In testing, the future is not always an issue. For example, achievement tests, or mastery tests, are designed to measure how well students are meeting the expectations of a particular instructional program, including those focused on language learning (Bachman 1990: 70). Achievement tests are not intended as predictions of the future but are measures of the person's history as a learner. In other words, they bring out actual rather than potential development. However, testing is also commonly used to make predictions about the future (and as Shohamy 2001, Hanson 1993, and others have argued, this is often for gate-keeping purposes), and when this occurs, past-to-present models of development are typically employed. For instance, aptitude tests construe the future based on present performance, essentially making a bet in favor of some individuals and against others. Similarly, proficiency tests, including language proficiency tests

such as the ACTFL-OPF, ostensibly predict (i.e., generalize to) future performance in non-test situations. Proficiency tests assume that the future and the present are equivalent; that is, future performance is taken to be a close reproduction of actual performance on the test itself.

Present-to-future models, on the other hand, concentrate on the *future-in-the-making*. Focus here is on the emergence of novelty, and for Valsiner (p. 86), the "research orientation of semiotic mediation belongs to the realm of these models." Not only do these models allow us to chart out development before it happens, "through their study *while* [italics in original] they are emerging" (this is what "proximal" means in the ZPD), they also compel us to participate actively in the developmental process itself. Researchers, and we would add, testers and teachers (arguably the same people) are concerned with the "process of the present (actuality), on the basis of anticipation of immediate future possibilities and through construction of reality out of these anticipated possibilities" (p. 86). By present, or actual development, Valsiner, echoing Vygotksy, means the person's past as it is brought into contact with the future.

Present-to-future models, then, predict the future not a priori but only on the basis of human agents performing in conjunction, or to use Vygotksy's term, cooperation, with other human agents. Paradigm examples of such models in psychology include Luria's (1976) research on cognitive development among the Uzbek peasant communities of the former USSR carried out in the early 1930s, Scribner and Cole's (1981) study of the differential effects of literacy among the Vai people of Liberia and Tulviste's (1991) replications of both studies among contemporary peasant communities of Russia. In each case it was found that participants who had Western schooling differed markedly in their cognitive functioning (e.g., completing basic syllogisms) from unschooled members of the community. This research shows that the classic Piagetian stages of cognitive development are not culture-free; rather than being the same for all human beings, cognitive development depends very much on the types of mediation people have access to as they go about the activity of living.

For testing, the future-in-the-making perspective sees ability not as a stable trait of an individual but as a malleable feature of the individual and the activities in which the individual participates. Thus performance on an aptitude test of any type, including language, is not complete until we observe how the person behaves in response to assistance. In other words, to fully understand the person's potential to develop (i.e., his/her future), it is necessary to discover his/her ZPD. Importantly, while we are gaining a perspective on the person's future, we are at the same time helping the person attain that future. Similarly, with regard to language proficiency, Lantolf and Frawley (1988: 188) argue that proficiency is not a property of an individual but is a feature of the functional system (see Luria 1973: 29 for a definition and illustration) formed between individuals and their unique, as well as shared, goals. In essence, their claim is that as with aptitude, proficiency is dyadic and is thus a feature of the dialogue that emerges between individuals (e.g., examiner and examinee) rather than a trait located exclusively inside of the person. Swain (2001) also makes the case for a dyadic approach to language proficiency testing. Citing the work of Lumley and Brown (1996), she points out that the linguistic features of an examiner's behavior during a proficiency interview can "differentially support or handicap a test candidate's performance", which means "performance is not solo performance, but rests on a joint construction by the participating individuals" (p. 278). McNamara (1997: 449) urges proficiency testers to abandon the assumption that proficiency is the cognitive activity of a lone individual (i.e., the examinee) functioning in a "curious kind of isolation" (p. 449) and proposes that proficiency tests need to take on a dynamic aspect so that "differences between learners and their potential for growth [i.e. future development] in the presence of assistance" should become part of both the testing procedure, and the rating scale (p. 454).

DA is very much in line with Valsiner's future-in-the-making model, since it is anticipated that future performance will be different from current performance. In essence DA sees the future

as a bet in favor of everyone. Lidz (2003: 103) stresses this point in the following quote: "traditional standardized assessment follows the child's cognitive performance to the point of 'failure' in independent functioning, whereas DA in the Vygotksian tradition leads the child to the point of achievement of success in joint or shared activity." Indeed, Feuerstein, Rand and Rynder's (1988) book on the Mediated Learning Experience, a form of DA, carries the revealing title, *Don't Accept Me as I am. Helping Retarded Performers Excel*.

In DA, as called for in Vygotksy's ZPD, assessment and instruction are dialectically integrated as the means to move toward an always emergent (i.e., dynamic) future. Bronfenbrenner (1977: 528), captures this notion very nicely in citing an excerpt from a conversation with A. N. Leont'ev, an influential colleague of Vygotksy, in which the latter scholar noted that "American researchers are constantly seeking to discover how the child came to be what he is; we in the USSR are striving to discover not how the child came to be what he is, but how he can become what he not yet is."

Methodological Differences

Given the differences in how SA and DA construe the future, it seems reasonable to assume that the two approaches would exhibit clear differences in how each concretely implements assessment procedures. Sternberg and Grigorenko (2002) enumerate three important and interrelated methodological differences between SA and DA. First, SA focuses on the product of past development while DA foregrounds future development. Said in Vygotksy's terms, SA taps into already matured abilities but DA promotes functions that are maturing.

Another difference between SA and DA has to do with the relationship between the examiner and the examinee. In SA examiners are expected to adopt a neutral and disinterested stance as a means of minimizing measurement error (Sternberg & Grigorenko 2002: 29). In DA the examiner-examinee relationship is markedly different in that the examiner intervenes in the assessment process and the "conventional attitude of neutrality is thus replaced by an atmosphere of teaching and helping" (ibid.). This leads to the final and crucial difference between SA and DA — the provision of feedback and mediation. In SA examinees are given sequences of problems to solve or tasks to perform and little or no feedback is provided on the quality of the performance until after the assessment is complete. To do otherwise would introduce a measurement error into the assessment. In DA, a very specific form of feedback is provided — mediated assistance — and this is the crux of the assessment process.

According to Sternberg and Grigorenko (2002: 29) mediation is offered in one of two formats — *sandwich* or *cake*?. The *sandwich* format follows a pretest-instruction-posttest procedure and can be administered in either an individual or group setting. If the individualized procedure is used, mediation can be implicit or explicit depending on the responsiveness of the examinee. In a group setting, mediation is usually implicit with respect to individual examinees because the examiner will be unable to inform each examinee of his or her specific strengths and weaknesses. The *cake* format lends itself to individual administration in that the examinee is provided with instruction during the assessment process itself. Mediation itself can be either standardized or individualized in DA, and this represents an important difference between a psychometric and a clinical orientation to DA (Lunt 1993: 164; Sternberg 2000: xv). This distinction will be fleshed out in the next section.

Before moving on, we must emphasize at this point that what makes a procedure dynamic or static is not the instrument itself but whether or not an intervention is incorporated into the process, regardless of where in the process the intervention occurs. In other words, fill-in-the-blank, multiple-choice, open-ended essay, or even oral proficiency tests are in themselves neither

static nor dynamic instruments. Their status is determined by the goal of the procedure and the format in which it is subsequently administered. For example, the ACTFL-OPI follows a static procedure in which the examiner leads the examinee through a series of question and answer exchanges, role plays, prompts, etc. but does not intervene to provide feedback or assistance to the examinee. As Antoni's (2003) work, discussed below, illustrates, oral proficiency can be assessed dynamically.

Interventionist vs. Interactionist Approaches to Dynamic Assessment

In this section we elaborate on the differences between the psychometric and clinical approaches to DA. The former we will refer to as *interventionist* and the latter as *interactionist*. The difference is reflected in Sternberg and Grigorenko's (2002) distinction between dynamic *assessment* and dynamic *testing*. They assert, "In essence the goal of dynamic assessment is to intervene and to change. The goal of dynamic testing, however, is much more modest - it is to see whether and how the participant will change if an opportunity is provided" (2002: 30). This is a significant distinction. Dynamic testing emphasizes the individual's or group's uptake of a predetermined repertoire of mediational means and as such attempts to discover the extent to which people will or will not change when offered pre-fabricated assistance. DA, on the other hand, seeks to promote development and therefore can neither limit nor pre-specify the types of mediation required and must therefore allow the appropriate assistance to emerge in the dialogue between examiner and examinee as they jointly engage in concrete tasks.³

Interventionist DA

Interventionist perspectives tend to follow a quantitative approach, and so lends themselves more to a psychometric orientation. Two well known interventionist approaches are Budoff's "learning potential assessment" and Brown's "graduated prompt" approach. We briefly consider each of these for illustrative purposes.

Using a pretest-intervention-posttest format, Budoff (Budoff 1987a & b, Budoff & Friedman 1964, Corman & Budoff 1973, Budoff 1968) developed dynamic procedures for administering several widely recognized static tests of mental ability, including the Raven Learning Potential Test, the Wechsler Intelligence Scale for Children, and the Wechsler Adult Intelligence Scale. Budoff argues that so-called general intelligence, or *g*, is trainable because it is influenced by the individual's sociocultural environment. His work focuses on individuals from at risk backgrounds "whose abilities were likely to be underestimated by traditional [static] IQ tests - specifically minority-group children and those from non-English-dominant homes" (Lidz 1991: 22).

Budoff's administers the pretest in a static format. An intervention stage follows in which the examiner takes on a non-traditional role, and through a standardized set of explanations helps the examinees understand the "crucial aspects of the task and the testing procedures" and guides them toward the correct solution (Sternberg and Grigorenko 2002: 75). The intervention can be administered in either a group or individual format. The next step is to readminister the original test at two different time intervals - one day and one month after the intervention. For Budoff highly standardized intervention strategies are needed in order to address what he considers to be a shortcoming of interactionist approaches such as Feuerstein's (discussed below), where "it is difficult to distinguish the contribution the tester makes to student responses from what the student actually understands and can apply" (Budoff 1987: 56). Budoff seems to be trying to determine how much of the performance can be attributed to the "environment" as represented by the tester and how much is to be attributed to the student. This, in our view, is a clear

parting of the ways with how Vygotsky theorized the person-environment relationship, as seen in the following:

One of the major impediments to the theoretical and practical study of child development is the incorrect solution of the problem of the environment and its role in the dynamics of age, when the environment is considered as something outside with respect to the child, as a circumstance of development, as an aggregate of object conditions existing without reference to the child and affecting him by the very fact of their existence. The understanding of the environment that developed in biology as applied to evolution of animal species must not be transferred to the teaching on child development. (Vygotsky 1998: 198)

In other words, for Vygotsky interaction between the child and the adult is "not a factor of development, not what acts from outside on what is already there, but a source [italics added] of development" (Elkonin 1998: 299).

The research of Brown (see, Brown and Ferrara 1985; Campione, Brown, Ferrara & Bryant 1984) extends DA beyond the domain of general intelligence to include specific content areas such as reading and math, arguing for domain-specific ZPDs. Thus, an individual may be able to perform math tasks with very little assistance while at the same time requiring a great deal of help to carry out reading activities.

In their procedure Brown and her colleagues first teach examinees to solve problems that require them to discover and apply a specific set of principles. Once these are mastered so that the examinees can solve the problems independently, the researchers then attempt to discover how far the individuals can transfer their new ability to novel problems. In each of the succeeding stages, the examiner provides assistance as needed from a pre-established menu of hints arranged from general to specific culminating with the solution to the problem. (In the initial posttest the examinees are given "novel exemplars" of the original problem types (Campione et al 1984: 81). Next they are given a set of "near transfer" problems which integrate the same principles in the original task but in new combinations. Then the examinees are presented with a set of "far transfer" problems requiring "the use of a new but related rule or principle in addition to the familiar ones" (ibid.). Finally, the examinees are asked to respond to a set of "very far transfer" problems that are even more complex. Based on the examinee's assisted and unassisted performance throughout the posttests, the researchers generate learner profiles comprising two axes - one measuring how quickly they are able to learn the new patterns and the other measuring how far they can extend this knowledge to novel problems (see Brown and Ferrara 1985).

According to Brown and her colleagues the graduated-prompt approach draws inspiration from Vygotsky in highlighting the importance of "interactive learning situations that provide structured guidance for the learner" (Campione, et al 1984: 80). They part company with Vygotsky in emphasizing the "metric of learning efficiency" (p. 82) as determined by "the number of hints required for the attainment of the learning criterion" (ibid.); whereas Vygotsky's focus is on how much development can be promoted through assisted performance. Here we observe a fundamental difference between interventionist and interactionist orientations to DA. Interventionist are concerned with quantifying as an "index of speed of learning" (Brown & Ferrara 1985: 300) the amount of help required for a learner to quickly and efficiently reach a prespecified end point. Using a train metaphor, Elkonin (1998: 300) states that those interested in speed and efficiency of learning are concerned with how quickly a train moves toward the final station along a set of tracks. Interactionists, on the other hand, follow Vygotsky more closely and are not so interested in the speed of the train along the already constructed track but with helping the person lay down new track leading toward a station that is potentially always being relocated (see Newman and Holzman 1993 on development as creativity and transformation). We flesh out the interactionist approach in more detail below.

Interactionist DA

Mimick (1987: 127) points out that for Vygotsky the ZPD is neither a way to assess learning potential (Budoff's interpretation of DA), nor a means of measuring learning efficiency (Brown and colleagues' understanding of DA), but "a means of gaining insight into the kinds of psychological processes that the child might be capable of in the next or proximal phase of development and a means of identifying the kinds of instruction, or assistance that will be required if the child is to realize these potentials." Unlike in interventionist orientations to DA, which have a strong propensity toward quantification and psychometric analysis, interactionist approaches follow Vygotsky's preference for "qualitative assessment of psychological processes and dynamics of their development" (Mimick 1987: 119). Indeed, Vygotsky (1988: 204) himself insisted that "we must not measure the child, we must interpret the child" and this can only be achieved through interaction and cooperation with the child.

Reuven Feuerstein, a leading advocate of interactionist DA, has produced a robust set of theoretical and empirical studies promoting a qualitative approach to DA that is very much in line with Vygotsky's understanding of the ZPD (Feuerstein, Kand, and Hoffman 1979; Feuerstein, Rand, Hoffman, and Miller 1980; Feuerstein, Rand, and Rynders 1988). Feuerstein, Rand and Hoffman (1979) argue that traditional conceptualizations of the examiner/examinee roles should be abandoned in favor of a teacher-student relationship in which both are working toward the ultimate success of the student. They write that "It is through this shift in roles that we find both the examiner and the examinee bowed over the same task, engaged in a common quest for mastery of the material" (p. 102). In this way, they attempt to bring instruction to center stage and to downplay the importance of psychometric measurements.

At the heart of Feuerstein's approach is the "Mediated Learning Experience (MLE)" – a construct reflecting Vygotsky's (1978) understanding of mediation. Feuerstein describes the MLE as a process through which environmental stimuli do not impact directly on the organism but are filtered through some other person, usually an adult mediator, who selects, frames, modifies, and imposes order on the stimuli to ensure that "the relations between certain stimuli will be experienced in a certain way" (Feuerstein, Rand, and Rynders 1988: 56). Sternberg and Grigorenko (2002: 54) point out that the mediator not only modifies the stimuli or task but also affects the learner by "arousing him or her to a higher level of curiosity and to a level at which structural cognitive changes can occur." For example, an adult watching a TV program with a child may mediate the program by explaining what the child is seeing. Like Vygotsky, Feuerstein understands mediation as "the psychological component of cultural transmission" (Feuerstein, et al 1981: 271). Unlike Vygotsky, however, Feuerstein assumes that some types of learning are non-mediated or direct, as for example when a child is watching a TV program alone.⁴

The MLE encompasses several important components, including feelings of competence, ability to self-regulate, and the internalization of general learning principles that guide the individual in "learning how to learn". In addition, the mediator must carefully select, schedule, and repeat as necessary, culturally determined stimuli for presentation to the individual. This enables the learner to more easily internalize the cultural practice he or she is participating in with the mediator. This process of internalization occurs as a result of the child's imitation of the models provided by the mediator⁵. In order to extend current abilities to future performance the individual must extend what has been internalized by anticipating outcomes that are likely to result from specific actions (Sternberg and Grigorenko 2002: 50-51).

The components of MLE are concretized into a dynamic procedure known as the LPAD (Learning Potential Assessment Device) that incorporates some of the well-known standardized

assessment instruments (e.g., Raven's Colored and Standard Progressive Matrices and the Rey-Osterrieth Complex Figure Test) with a mix of instruments designed specifically for the LPAD. The LPAD requires the examiner to interact flexibly with the individual examinee, negotiating the assistance and guidance required to "modify the cognitive structure of the individual" (Feuerstein, Rand & Rynders 1988: 204). The examiner thus functions as a mediator who reacts to the learner's responsiveness and is more concerned with cognitive transformation than with performance efficiency.

In Mimick's (1987:138) view, Feuerstein's model reflects Vygotsky's ZPD in allowing the assessor greater freedom to interact with the learner and thereby deploy a wide array of assistance to help the individual develop. In addition, the LPAD allows for a true diagnostic assessment because it brings to the surface the psychological processes underlying learner performance. One problem with the reporting of Feuerstein's research, as noted by Mimick, is that it fails to provide extensive data on these psychological processes. However, two recent interactionist studies, Karpov and Gindis (2000) and Peña and Gillam (2000), document the processes in considerable detail.

Karpov and Gindis (2000) report on a set of case studies of analogical reasoning in children with learning problems. One study focused on a seven year-old child who was diagnosed as immature and as having limited cognitive and linguistic abilities and attention-deficit-hyperactivity disorder (ADHD). During a pretest, the child was unable to reason even at the visual-motor level (e.g., count objects by touching or moving them). At the outset of the DA procedure, it took a good deal of time and effort for the evaluator to focus the child's attention on the explanation and manipulation of geometric shapes (a square, a circle, a triangle, a cross, and a star) required to carry out the reasoning task. In talk, the evaluator succeeded in focusing the child's attention on the task, and eventually the child began to show an ability to reason at the visual-motor level (i.e., could complete an analogy representing the relationship among a series of geometric shapes by physically manipulating the shapes). The evaluator then pushed the child to the next level (i.e., visual-imagery reasoning) that required the child to complete similar analogies by only looking at the shapes rather than physically moving them. Finally, the evaluator ceded increasing responsibility for solving the problems to the child by encouraging her to use private speech to regulate herself during task performance. Through DA it was shown that contrary to the results of a SA procedure, the child was not cognitively deficient but was in fact able to perform at an age-appropriate level once she learned to use her own speech for self-mediation, thereby overcoming the challenges presented by ADHD (p. 151).

Peña and Gillam (2000) discuss a series of DA case studies on distinguishing children with language impairment ("unusual difficulties learning language," p. 543) from those with problems arising from language differences (e.g., bilinguals, nonstandard dialects, etc.). Focusing on three domains of language (vocabulary, narrative ability, and discourse performance), the researchers rely on individualized and highly interactive mediation and follow a pretest-mediation-posttest format. For instance, some of the children often have difficulty understanding the importance of using single words to denote objects, events, and concepts. Working on this problem, Peña and Gillam ask the children leading questions such as, "Have you ever known someone who was _____?" and "What does it mean when X said Y?" They also ask more open-ended types of questions such as, "What would happen if the puzzles were moved to the art area?" (p. 553).

In one study, the performance of a four year-old Spanish-English bilingual child on the *Expressive One-Word Picture Vocabulary Test-Revised (EOWPVT-R)* was below normal, but on the basis of her performance alone it was not possible to tell whether this was due to the linguistic and cultural bias of the test or to a genuine language impairment (p. 551). For most test items, she was either non-responsive or simply replied, "I don't know." Through a DA procedure, Peña and Gillam were able not only to uncover the source of the child's problem but also to provide media-

tion to help her overcome the problem to some extent. While her performance on the EOWPVT-R did not improve following mediation, she did show improvement in her ability to self-regulate and plan, as well as in her motivation and attention to the task. Based on the DA, the researchers concluded that the child was suffering from a language impairment and did not just have a language difference problem. They also made a series of recommendations the teacher could implement in the classroom setting to help the child develop her vocabulary despite the impairment.

L2 Dynamic Assessment

In this section we will review the only five L2 studies we are aware of that are carried out from a DA perspective. Admittedly, this research, still in its nascent stage, has yet to explore the full potential that DA can bring to the field. Two of these (Guthke, Heinrich and Caruso 1986 and Kozulin and Garb 2002) we situate within the interactionist framework, and two (Schneider and Ganschow 2006; Antón 2003) we see as more interactionist in their orientation. The fifth study (Grigorenko, Sternberg and Ehrman 2000) positions itself within the DA tradition, but for reasons that will become clear, we do not agree that it belongs to this paradigm and we will dispense with this study first.⁶

Grigorenko, Sternberg and Ehrman (2000) report on a foreign language aptitude study which they claim is based on DA principles. The authors argue that their framework, unlike previous aptitude tests, is based on a theory of learning that they refer to as CANAL-F (Cognitive Ability for Novelty in Acquisition of Language (Foreign)). The theory maintains that successful foreign language learning is predicated on the ability to cope with novelty and ambiguity (p. 392). This is operationalized in an assessment instrument called the CANAL-FT, comprised of nine subsections related to various components of the authors' theory of aptitude. Examinees are presented with an invented language, Urusulu, and are led through a series of graded tasks relating to its lexicon, morphology, semantics, and syntax.

According to the authors, the CANAL-FT is a DA instrument because it taps "the processes of knowledge acquisition at the time of the test" (p. 393). Specifically, the lexical, morphological and syntactic features of the language are gradually introduced as examinees move their way through the subsections of the test. Nowhere during the administration of the CANAL-FT, however, does a mediator intercede, and this, after all, is the defining quality of DA. The fact that the test presumably taps the examinee's ability to learn during the test is not what makes a test dynamic; rather it is the provision of feedback and assistance by the examiner that marks a procedure as dynamic (see Sternberg and Grigorenko's 2002 definition provided above). CANAL-FT privileges solo performance and for this reason we do not consider it to be legitimate DA.

Interactionist Approaches to L2 DA

Guthke and his colleagues conducted a language-based version of the *Lernitest*. In the procedure items are presented in a sequence of increasing complexity; learners are provided with immediate implicit or explicit feedback as needed until they are able to respond appropriately to an item (i.e., the cake model), the assumption being that they cannot move to more complex items until they have mastered the principles underlying simpler problems (Guthke 1982: 316).

The extension of the *Lernitest* to foreign languages is aimed at assessing the aptitude of international students required to learn German prior to enrolling in German universities. Guthke, Heinrich and Caruso (1986) incorporated a subsection of the Carroll MLA test into a *Lernitest* procedure. Similar to the CANAL-FT, the *Lernitest* presents examinees with the challenge of discovering the rules and principles of an invented language, but with the crucial addition of immediate

feedback from the examiner. Examinees are given sets of geometric figures paired with words from an invented language and are asked first to memorize the associations (i.e., the equivalent of referential meanings of words) and then to figure out the meaning of a given sequence of symbols. If the examinee fails to produce the appropriate meaning they are first reminded "to think properly once again" (p. 905), and if this fails the examiner provides the appropriate response and at the same time presents the examinee with the next problem. Learners have full access to all of the earlier items as they work their way through the test. This not only lessens the memory burden but also provides continuous feedback.* *Lernitest* scores reflect the time taken to complete the test and the total number of prompts required as well as changes in the number of prompts needed during the course of the test. Thus, an individual requiring the same number of prompts on later items as on earlier items is given a different score from a person requiring fewer prompts toward the end of the test than at the beginning, which would be interpreted as an indication that such an individual learned through the test.

Kozulin and Garb (2002) report on a subset of a larger study on EFL reading comprehension skills conducted among young adult at-risk immigrants to Israel. The authors developed an instructional curriculum that included a DA component focusing on helping learners develop general comprehension strategies that could be used to access meaning in a broad array of texts, regardless of a given text's vocabulary and grammatical properties. Students were first asked to read a simple passage in English and to answer a set of comprehension questions. Following the pretest classroom teachers, who had been trained as mediators, reviewed the test with the students, "mediating for them the strategies required in each item, building together with the students process models for each item, and indicating how strategies can be transferred from one task to another" (p. 119). Additional forms of mediation included 'an information page' that was given to students which included grammatical and lexical information relevant to the pretest text (e.g., use of auxiliary verbs 'to be' and 'to do'; question words, formation of negation, etc.) and an activity that required students to read four texts and answer comprehension questions focusing on text structure, cohesive devices and background knowledge. Unfortunately the authors do not specify precisely how the teachers mediated the test items, whether this was done individually or in a group-format, and what the focus was (e.g., grammar, text structure, etc.). Following the mediation students completed a posttest that closely paralleled the pretest.

Kozulin and Garb argue that the students' abilities could not have been fully captured on the basis of their performance on the pretest alone; rather it was necessary to determine the extent to which they had benefited from the intervention. They therefore devised a formula to calculate what they call the Learning Potential Score or LPS, which quantified the gain between the pre- and post-tests. This, they argued, provided a more complete picture of the learners' ability because it allowed the researchers to group students as low, moderate and high performers, and to make instructional recommendations for how to better help each of the groups in the future.

Interactionist Approaches to L2 DA

Turning to the interactionist DA work on L2 assessment, we first consider a publication by Schneider and Ganschow (2000) – a 'think piece' rather than an empirical study – that speculates on the potential of DA in helping at-risk second and foreign language learners, in particular, those with problems arising from dyslexia. Schneider and Ganschow suggest using DA procedures to help at-risk language learners develop metalinguistic awareness, which they believe will facilitate learning. According to the authors, such awareness is important, particularly given what they see as the strong penchant in L2 pedagogy for communicative language teaching (CLT), which they

argue favors implicit rather than explicit learning of language features. Although we do not necessarily agree with their take on CLT, we appreciate their argument that explicit instruction of the L2 system is helpful for at-risk learners. Indeed, the research of these authors and some of their colleagues provides empirical support for this claim (see Schneider 1999; Sparks and Ganschow 1993a and b).

Following the work of Baker and Brown (1984), the authors outline two types of metalinguistic awareness—knowledge of the linguistic rule system and knowledge of strategies for applying the metacognitive system. They envision a classroom-based DA in which L2 learners are helped, through interactive dialogue, to develop knowledge and strategies for application of knowledge that are required during L2 performance. In an “assessment cycle” the teacher serves as both a facilitator of knowledge and a collector of data on the learner’s metalinguistic awareness as it evolves over time in order to pinpoint sources of difficulty. The students, for their part, learn to become active self-monitors in planning and evaluating their performance as well as to function as reciprocal respondents who consistently supply the teacher with information on the reasons for choices made during specific performances. In this way, teacher and student enter into a mutually dependent and dialogically mediated relationship with the teacher assessing and promoting progress and the student controlling its speed and direction (p. 76).

A conference presentation by Antón (2003) also reports on work done on L2 interactionist DA. The study documents the use of interactive DA procedures to place students in university advanced Spanish language classes. Student performance was evaluated on the basis of accuracy in the use of sentence-level grammar and vocabulary. The goal of mediation in Antón’s study was not so much to promote learner development as it was to generate a diagnostic evaluation that would inform placement decisions so that students could enter courses where they would receive more finely-tuned instruction. In what follows we consider two representative excerpts from Antón’s protocols in which the examiner prompts the students at various points in order to give them the opportunity to revise their performance in appropriate ways. In terms of the evaluation and placement, students who were able to revise under prompting were considered to be at a more advanced stage than students who could not. We present the English translations of the protocols and give the Spanish only where it cannot be avoided.

In the first example, a student attempts to use Spanish past tense verb forms while relating an oral narrative based on a short film about a family traveling through Spain. Upon completing the narration the examiner asks the student some questions about details of the story. As the student is responding, the examiner interrupts and comments on the linguistic quality of the performance. We pick up the exchange at this point:

- (1)
 (Examiner: You started the story in the past and then, half way you switched
 (Student: Yes, yes
 E: To the present.
 S: Yes, yes. I heard
 E: Do you want to try again using the past? And you can ask me. If there is a verb you do not remember it’s OK.
 S: Yes, yes, from the beginning?
 E: Perhaps from the middle
 S: In the past, yes, yes.
 E: Did you realize that you made the switch?
 S: Yes, yes, I heard.

The student then renarrates the story from the middle and, with only two or three slips back into the present, uses the appropriate past tense forms.

What is evident in the preceding protocol is not only that the student was indeed able to appropriately renarrate the story when given the opportunity but also that he was aware of the problem even before the examiner overtly mentioned it. In a static procedure, the student most probably would have been given a lower evaluation since he appeared to be unable to sustain the narrative in the past tense. This clearly would have been an inappropriate assessment, since the learner was able to complete the task as required when given the chance to do so. In terms of development, we agree with Antón that the learner had a greater degree of control over past tense than a static test would have indicated. To put it in Vygotsky’s terms, the past tense was in the process of maturing and the learner required only a leading comment to make this manifest.

In a second protocol, another student, asked to narrate the same film, used the present tense exclusively throughout the narrative. When given the opportunity to renarrate the story, this student attempted to comply but had difficulties marking appropriate person features (using first person instead of third person forms) and could only do so when given the choice between two options by the examiner. In addition, when the student renarrated he encountered a lexical problem which he again solved with help from the examiner. We pick up the narrative from this point.

- (2)
 S: She ... arrived at the wall of the bus and ... waited with her friends at the wall
 [Here the student uses the Spanish word pared ‘wall’ instead of the appropriate para-da ‘stop’]
 E: Wall or stop?
 S: Pared o parada?
 E: Stop
 S: Parada
 E: Do you know what pared is?
 S: ... wall
 E: It’s a very similar word, isn’t it?

They then returned to the narrative, but shortly thereafter the student began to have problems with the past tense. We continue the protocol at this point:

- (3)
 S: I played tennis
 [Jugué al tenis
 E: I played or she played?
 S: She played
 [Jugó]

A bit later a similar problem arose when the student was attempting to narrate the fact that one of the characters returned home to eat lunch.

- (4)
 E: Very good. And here you said, what did she do?
 S: I ate
 E: Comí
 S: I ate or she ate?
 E: Comí o comió?
 S: She ate
 E: Comió
 S: She ate

In excerpts (3) and (4) the student was unable to produce the correct verb form without assistance and, importantly, did not seem to appreciate the assistance since following the interaction in (3) he required the same type of explicit feedback in (4). From an SA perspective both this student and the student presented in protocol (1) had problems with past tense. From their respective interactions with the examiner, however, it is clear that the learners have different levels of control over the relevant features, as shown by their need and responsiveness to different types of feedback. Therefore, they have different ZPDs for the same feature and consequently were placed in different classes.

Criticisms of DA: The Psychometric Turn

Despite a robust research literature, DA has not been accepted with open arms by the educational and psychological testing communities. The lion's share of criticism leveled against DA by those who "are comfortable and familiar with static tests" (Sternberg 2000: xiv) stems from perceptions of psychometric shortcomings. In this section we address some of these issues.

Some critics of DA such as Snow (1990) do not accept the dynamic-static contrast proposed by DA researchers. Snow dismisses the distinction as a "propaganda device" (p. 1134). According to Snow, the term 'static' is a misnomer because in his view 'static' tests, like 'dynamic' tests, are concerned with predictive validity (i.e., the future). We do not disagree with Snow on this point, but we point out once again that SA and DA differ with regard to the nature of the prediction they make. As argued above, in DA prediction is about an imagined future that emerges only through mediated activity (i.e., genuine development); in SA, prediction is about generalizing an individual's test performance to subsequent non-test situations. In other words, to recall Valsiner's (2001) discussion, DA predicts a future-in-the-making while SA predicts a future-in-the-present. Therefore, we do not agree with Snow's accusation that the distinction between SA and DA is mere propaganda.

Snow also objects to using the term 'assessment' to describe DA procedures. He argues that without linking assessment in some way to measurement, "fundamental in all science," the term is "meaningless" (p. 1135). Indeed, concern over measurement is the crux of the psychometricians' discomfort with especially interactionist DA. In language testing, Bachman has similarly suggested that terms such as 'assessment' and 'appraisal' are little more than "stylistic variants" of 'evaluation' and 'test', and in his view their use reflects 'impressionist [sic] approaches to measurement' (Bachman 1990: 50-51, note 1). For Bachman (1990: 18), measurement is "the process of quantifying the characteristics [physical as well as mental] of persons according to explicit procedures and rules." One way of achieving this in DA, according to Büchel and Scharnhorst (1993: 101), is through "standardization of the examiner-subject interaction." On this view, interactionist approaches to DA would seem to fit the bill. On the other hand, the interactionist orientation to DA would not. Glutting and McDermott (1990: 300), consequently, assail Feuerstein for relying on what they see as improvisation and "creative latitude in the administration of mediated learning experiences," and the fact that some children receive more help than others during the interaction is for them problematic. Finally, they worry about 'instrument decay' as "teaching prompts on early test items can result in autonomous changes in the difficulty level of subsequent items" (*ibid.*).

For Vygotsky, improvisation and creativity are essential to providing appropriate forms of mediation in the ZPD (Newman and Holzman 1993), while measuring a child's performance provides little more than "a purely empirical establishment of what is obvious to persons who just observe the child" and adds nothing new to what is already known through direct observa-

tion (Vygotsky 1998: 205). The task of the psychologist is not to measure but to interpret the child (Vygotsky 1998: 204), and this is what is meant by assessment.

Reliability

Another criticism of DA concerns test reliability and standardization; presumably, without standardization there can be no reliability (Büchel & Scharnhorst 1993: 103). Traditionally, test reliability derives from a commitment to standardization whereby all sources of potential error should be minimized to ensure that the observed score on a test is as close to the true score as possible. Reliability assumes that what is being measured is more or less stable. Within DA, interactionist researchers, while not viewing traits as stable, have nonetheless attempted to reduce measurement error through reliance on standardized forms of assistance. Interactionist approaches, on the other hand, are more problematic when placed under the psychometrician's lens. As with their interactionist counterparts, interactionist assessors argue that abilities are inherently unstable, but they further argue that to be maximally useful in promoting development, assistance must be tailored to the needs and responsiveness of the individual learners, a requirement which undermines standardization. In Feuerstein's approach, for example, "everything is done in order to undo the predictive value of the initial assessment by modifying functioning through the mediational process" (Feuerstein, Rand, and Rynders 1988: 199). In what to some is no doubt a curious turn, the more reliable the procedure, the less effective it is in promoting individual development. As Lidz (1991: 18) cogently puts it, "the word 'dynamic' implies change and not stability. Items on traditional measures are *deliberately* selected to maximize stability, not necessarily to provide an accurate reflection of stability or change in the 'real' world."

Even in mainstream assessment, as less standardized approaches (e.g. performance assessment) have gained acceptance, reliability has become more problematic because such assessment instruments imbue examinees with greater "latitude in interpreting, responding to, and perhaps designing tasks" leading to different performance outcomes (Moss 1994: 6). Swain (1993), in discussing the variable performance of French immersion students on a series of oral as well as on a series of written tests, questions the psychometric assumption that for a test to be appropriate it must have high internal consistency. Indeed, citing the work of Shohamy (1988), which shows that reading ability varies across genre, register, topic, etc., Swain (1993: 202) notes that "One might wish to argue that a good test of second language reading proficiency must have a low internal consistency." A recent study by Spence-Brown (2003) on embedded performance assessment (see below for details) of Australian learners of L2 Japanese uncovered considerable variation in learners' interpretation and execution of the task.

Validity

While reliability is detrimental to interactionist DA, validity is not. As mentioned earlier, DA derives its validity not from the specific instruments used in the assessment process but from the procedures followed in the administration of the assessment instrument. Given that the purpose of DA is to push the person's (language) abilities forward, to the extent that this is achieved, the validity of the procedure is established. Of course, researchers may differ with regard to how they understand the abilities they want to assess (e.g. is proficiency the ability to use language effectively to achieve some goal or is it the ability to use language fluently and accurately in terms of grammar, lexis, pragmatics, pronunciation etc. or is it all of these?). Nevertheless, DA can be used to promote development in any or all of these dimensions, given appropriate mediation.

In discussing the work of Fredricksen and Collins (1989) on systemic validity (see below), Shohamy (2001: 142) argues that a test is valid if it "brings about or induces, an improvement in the tested skills after a test has been in the educational system for a period of time." This may seem to be a position that mirrors DA, but in fact, it is not. The claim here is that information gleaned from a test, if properly integrated into a curriculum can have a positive impact on learning and teaching (ibid.). DA, on the other hand, argues that it is the assessment procedure itself that promotes learner development; development then is not something that happens exclusively after the assessment has been carried out.

Guthke and his colleagues have undertaken to psychometrically establish the construct validity of the Lerntest. The interactionist tradition, aligned with Vygotsky's understanding of the ZPD, establishes construct validity through a qualitative analysis of the psychological processes that underlie an individual's performance. Feuerstein, Rand and Rynders (1988: 205) state that in DA, "very little attention is given to product or to the absolute magnitude of a result. More importance is attached to learning about the process that has brought about a particular product."

Carlson and Wiedl (2000), echoing Messick's (1988) recommendations, argue that an equally important type of validity for DA is consequential validity, particularly with regard to "the context of and justification for its use, a result of pragmatic judgments combined with scientific analysis. If a measure cannot be justified for its practical utility it becomes irrelevant" (p. 708). For instance, in Antón's (2003) work consequential validity took center stage as the Spanish faculty were able to make better informed placement decisions for their advanced learners.

Generalizability

Another issue that has been raised in connection with DA by some L2 researchers is the matter of "generalizability" (Alan Davies, Merrill Swain, personal communication, Sept. 5, 2003). Generalizability can be interpreted in two ways, depending upon whether the assessment instrument is being used as a basis for making decisions about individuals or for research purposes. We have already addressed the first type of generalizability in our discussion of the future in DA and SA. Generalization for research purposes often involves drawing inferences about a general population from a representative sample of that population. We now turn our attention to this view of generalizability, particularly as it relates to DA research. As Kvale (1994) observes, "a demand for generalization has loomed heavily in the social sciences. To the critical question 'why generalize?' the answer would probably be: in order to predict and control, or because science aims at universal knowledge" (p. 164). He notes, however, that despite the "quest" for generalizability, research in the social sciences in this regard has produced "meager results" (ibid.).

Presumably one of the concerns of the psychometric stance on generalizability arises from the fact that DA, particularly interactionist approaches, strongly favor a one-on-one dialogic relationship between examiner and examinee and therefore are not likely to generate the kind of numbers expected in psychometric research. This, in itself, does not eliminate the importance of generalizability. Kvale (1994: 165) answers the question "how many interview subjects do I need?" with the response "interview so many subjects that you find out what you need to know ... new interviews are conducted to a point where further interviews yield little new knowledge, until the law of diminishing returns applies." Vygotsky (1997: 317) also points out that through intensive and close analysis of a single phenomenon or organism it is possible to "deduce a general regularity" [italics in original]. An even more powerful argument, which goes beyond sheer numbers, is proposed by Kvale (1994), and much earlier by Vygotsky (1997), to the effect that a well developed theory allows us to legitimately generalize the findings of particular cases if we

can connect the analysis and interpretation to the statements and principles of the theory. Polkinghorne (2003) refers to this process as the "assertive logic" typical of legal argumentation and analytic generalizations based on theory. This is what Minick (1987) has in mind when encouraging DA researchers to pay closer attention to connecting their work with Vygotsky's theoretical framework.

Revisiting the Teaching / Assessment Interface

Sternberg and Grigorenko (2002: viii-ix) note that testers have for some time been tinkering with the "cosmetic" aspects of tests, such as computerizing traditional paper and pencil tests, or conducting oral interviews in an on-line format. In their opinion, however, what is needed is a paradigm shift that integrates instruction and assessment as a unified activity. The field of language testing has recognized the need to bring testing into more intimate contact with acquisition research and pedagogical practice. Early work by Swain (1984: 195), for example, argues that in communicative language tests it is essential to "bias for best." Among other things, this means that tests should minimize "the effect of the measurement technique on the test-taker's performance" (ibid.) and further that the testee might even learn "something from the experience of being tested," particularly if the test contains new content (p. 193). While this is certainly a welcome perspective, it does not go as far as DA in unifying teaching and assessment into a single activity where the goal is necessarily development, and this depends not on the instrument itself but on the nature of the mediation provided to the individual.

In the introduction to their aptly titled volume, *Interfaces Between Second Language Acquisition and Language Testing Research*, Bachman and Cohen (1998) suggest that "since SLA tends to focus on the process of language acquisition, and LT on its product, the combining of qualitative and quantitative approaches would appear to provide a particularly powerful paradigm for future research in this interface" (p.9). They provide several examples of the "interplay" (p. 7) between SLA and LT. They mention, for instance, that research on interactional analysis led to more communicative tests of speaking ability and to an increased interest in assessing lexical knowledge required for adequate performance in school settings, and that SLA research on pragmatics led to the emergence of efforts to assess sociolinguistic and sociocultural abilities. The following subsections address two additional areas of research that relate to the teaching/testing interface - test washback and task-based assessment - and contrast these with DA.

Washback

To better understand the relationship between testing and instruction researchers have paid increasing attention to the "washback" effect of tests and the broader issue of systemic validity. Washback concerns the impact of tests on the curriculum and on teaching and learning practices (Swain 1984; Alderson & Wall 1993; Wall & Alderson 1993; Wall 2000). Tests are said to have negative washback when they narrow the curriculum and constrain teaching and learning practices and positive washback when they promote learning that extends beyond a particular test (Messick 1996). Fredricksen and Collins (1989) suggest that a test has high systemic validity (i.e. positive washback) if it promotes instruction and learning and low systemic validity (i.e. negative washback) if it results in a breakdown in the testing-instruction-learning cycle.

In our view, washback and systemic validity are issues for mainstream testing because of the well entrenched teaching/testing dualism that has dominated Western education and psychology for more than a century.⁷ Even in the writings of a leading proponent of the critical

language testing movement, we see evidence of the dualism at work. Shohamy (2001), in arguing for the need to convert tests from instruments of power to more democratic instruments of learning, raises the following question: "How can tests be used to maximize their role as a means for obtaining information relevant to the improvement of learning, while at the same time minimizing their power and control roles?" (p. 141). The framing of the question itself belies the underlying dualism. Tests are used for "obtaining information" that is then made available to students, teachers, and administrators for "the improvement of learning." Clearly, then, tests are temporally separated from curriculum building and teaching practices. To reiterate, DA is the simultaneous and dialectical integration of assessment and instruction. As such, these activities are two sides of the same coin, and there are no one-sided coins. Guthrie's *Lernfest* 'learning test' concept nicely captures this relationship. In SA, assessment and instruction are not only different coins, but they are often different currencies and therefore researchers continue to talk about these activities as related but independent enterprises.

Task-based assessment and DA

Researchers have also begun to draw out the implications of L2 task-based research and pedagogy for language assessment (see Skehan 2001, Chalhoub-Deville 2001 and Wigglesworth 2001). The advantages ascribed to task-based assessment include: it promotes the "test-taker's individual expression" and provides the opportunity for test-takers to use "background knowledge and experiences" as "active and autonomous participants in a given communicative interaction"; if one accepts that a task reflects real-life communication (a controversial issue as Spence-Brown's (2003) research clearly demonstrates), it is possible to generalize beyond the "learning/testing situation" (Chalhoub-Deville 2001: 214). Echoing the recent work of Bachman and Cohen (1998) Shohamy (2001) on testing in general, Chalhoub-Deville (2001: 224) argues that task-based assessment can inform pedagogical practices, but only if scores are accompanied by appropriate diagnostic information on the learners' abilities.

Although bringing tasks into the assessment picture is certainly a positive step, the task-based framework continues to separate assessment from instruction. Consider, for example, the Hong Kong primary school Target-Oriented Curriculum, as discussed in Candlin (2001). The TOC is organized around 'learning targets' and these in turn are operationalized in terms of concrete tasks. However, the curriculum segregates learning tasks from assessment tasks, as seen in the following excerpt taken from the curriculum: "the major difference between assessment tasks and learning tasks is that in learning tasks, teachers need to conduct appropriate pre-task, while-task and post-tasks activities to ensure that learners can complete the tasks satisfactorily" (Candlin 2001: 237). Thus, while a given task could be used for either learning purposes or assessment purposes, it is clear that these are still distinct activities.

Alternative Approaches to Assessment, Embedded Task-based Assessment and DA

McNamara (2001: 343) calls for an "alternative approach to assessment" which moves beyond so-called "alternative assessment" normally associated with portfolio, project-based and performance assessment. Alternative approaches to assessment attempt to bring instruction and assessment into more intimate contact. For one thing, teachers are not expected to compare performance across individuals (i.e., norm-referenced testing) nor are they interested in comparing

an individual performance to a "particular yardstick" (i.e., criterion-referenced testing) (ibid.). Furthermore, even the comparison of individual learner performance over time is carried out descriptively rather than through score comparisons, thus avoiding "the domain of measurement" (p. 344). McNamara points out that in this way, one can steer clear of many of the problems affiliated with so-called "subjective assessment" (ibid).

Despite the fact that this proposal maintains many of the features traditionally associated with assessment (e.g., samples of performance, reference to criteria, makes decisions about interventions, etc.), McNamara worries that it will be objected to as being about pedagogy and not about assessment (p. 344).⁸ McNamara acknowledges that he is blurring the boundary between pedagogy and assessment and while we find this to be a positive move, it still lacks the qualities of DA. McNamara, for example, states that in his approach the testing researcher's role would include investigating "effective procedures for promoting critical reflection [italics added] on learner performance by teachers" (p. 344). He also proposes that researchers should develop ways of helping learners to "engage in a process of deliberate sustained reflection [italics added] of the quality of the products of their learning to date" (p. 345). He further suggests joint evaluation of performances carried out by learners themselves as they collaborate on tasks such as dictoglosses and puzzles (see Swain 2001). While reflection is certainly an important part of development, as Vygotsky (1987) points out, alone it is not sufficient to promote development. Concrete goal-oriented activity is needed to push development forward (Vygotsky 1987), and this is precisely the aim of DA.

One way of concretizing McNamara's proposal for an alternative approach to assessment is explored in Spence-Brown's (2003) previously mentioned study of an embedded assessment task in which L2 learners of Japanese conducted interviews with members of the Japanese community in Melbourne in an extra-classroom setting. Embedded assessment is "the use of tasks which serve a pedagogical purpose for assessment" and has the advantage of providing robust feedback and allowing for the washback effect to motivate students to achieve specific short term goals (p. 36). One of the problems Spence-Brown sees with embedded assessment, however, is that the product of learning "does not represent what a student can do unassisted, after the learning cycle is completed, but what they can do with support during the learning process" (ibid.). In her view this is an advantage for formative assessment, but a threat to validity for summative assessment, which seeks to uncover what the learner can do independently after the learning cycle is completed. In DA, both assisted and unassisted performance are central to understanding and promoting learner development. Moreover, because DA is grounded in Vygotsky's future-in-the-making model, where development is emergent, the learning cycle is in principle never completed, other than in a trivial sense as when one completes a prescribed syllabus for a language course. Therefore, from a DA perspective the distinction between formative and summative assessment makes little sense.

A second problem Spence-Brown finds with embedded assessment relates to the washback effect on learning. According to Spence-Brown, because embedded assessment tasks are designed to serve the dual function of learning and assessment they are assumed to have positive washback on learning. However, for most of her participants the assessment quality of the interview task was overpowering in that "it made it difficult for most of the students to also frame it as an authentic interaction, and so this aspect of the task was greatly diminished" (p. 233). She considers this to be negative washback because the learning function of the task was subverted. For example, several of the learners produced pre-prepared questions, the answers to which they knew in advance. In this way they undermined the authenticity of the task as a conversational interaction. In fact, one learner rerecorded her interview because she had failed to use a clarification request as indicated in the teacher's instructions; in her rerecording, she followed the directions and asked for clarification of a word, even though she already knew its meaning. Spence-Brown

also sees the task as having positive washback effects. For example one student felt compelled to use honorifics in interacting with her interviewee because this was a requirement imposed by the teacher's instructions, something she would not have done on her own. Regardless of whether the task led to positive or negative washback, it seems clear that from the students' perspective the boundary between teaching and assessment was not blurred by this task as the learners were oriented not toward learning through the task but toward earning a good grade.

According to Spence-Brown, in carrying out the interviews, some of the learners were functioning within their ZPD while others were not. Her argument is that the learners' performance was mediated by pre-prepared interview questions that were formulated with the use of textbooks, dictionaries and teacher feedback. She points out that this allowed the learners to perform at a higher level of proficiency than they would have without such support. While we agree in part with Spence-Brown's interpretation, we see the situation as more complex when viewed from a DA perspective. There are two problems with her analysis. One is that there is no evidence of the learners' actual level of development in Japanese because all of the learners worked from pre-prepared scripts and we cannot properly assume that their scripts represented their actual developmental level, since they received feedback and assistance from the teacher as well as from pedagogical artifacts such as dictionaries, grammar books and the like. The second problem is that some of the students asked questions not contained in the script, and indeed one even abandoned the script almost completely at certain points in the interview. Although this is certainly a positive feature of the interviews, had the mediation been more flexible and tuned to the student's behavior in the concrete activity as it unfolded (i.e., the mediation was neither graded nor contingent nor interactive) they might have been able to perform at an even higher level of ability. In this way we would have had a sharper picture of their future development in the making. In addition, one student failed to deviate at all from the pre-prepared script, because he was overpowered by the testing orientation of the task and was afraid to take any risks that might result in a lower evaluation by the teacher. In this case, we have no way of knowing what the student's actual level of development was nor can we make any claims about his ZPD because his actual level of development was never pushed.

One way to conduct an embedded task-based assessment such as Spence-Brown's in a dynamic framework would be to ask the learners to first conduct the interview without prior preparation, then to provide feedback on their performance and recommendations for how to improve it, and finally to ask them to redo the interview, perhaps with the same individual but focusing on somewhat different topics, or with different individuals but with more or less the same topics. Another possibility would be to have the learners conduct their interviews in the presence of a more advanced peer, tutor, or teacher whose responsibility would be to provide appropriate assistance that is sensitive to the learner's responsiveness. This would then be followed by another interview carried out by the learner without assistance, and again varying the interviewee and/or topics. Once assessment and instruction are dialectically unified, as we suggest here, the boundary between them is not just blurred but eradicated.

Conclusion

Summarizing the distinction between SA and DA, the former traditionally isolates individuals from all forms of external mediation while the latter is specifically designed to provide such assistance. Furthermore, SA has been very concerned with test instruments so that scores on tests are taken to be measures of the amount of know/edge gained as a result of instruction, the amount of ability possessed in order to benefit from instruction, or the likelihood that what has been learned can be extended to non-assessment contexts. DA, on the other hand, focuses on promot-

ing change in the learners, either in terms of efficiency of learning, as in an interventionist model, or in the psychological processes underlying development, as in an interactionist approach. DA recognizes that mediation (a form of instruction) is a necessary feature of genuine assessment. SA foregrounds the test instrument while DA foregrounds individuals.

With regard to the interactionist-interventionist distinction within DA, it is clear that interactionist approaches are closer to Vygotsky's notion of the ZPD and how he understands the processes at work in human development. They therefore prefer a clinical perspective on diagnosing ZPDs and helping individuals develop. Interventionist approaches, on the other hand, focus on helping individuals become more efficient in their learning, and therefore have retained "some of the psychometric properties of more traditional 'static' forms of assessment" (Lunt 1993: 164). In our view whether one opts to use an interventionist or interactionist approach depends on the goal and circumstances under which an assessment is to be conducted. Interactionist approaches, because they are more labor-intensive and time-consuming, are likely to be more useful in a classroom setting involving relatively fewer students. Whereas for dealing with larger populations of learners the interventionist paradigm seems more appropriate. The tradeoff is that in gaining efficiency, one loses access to unique information on psychological processes that can only be brought out through interaction with individuals (Lunt 1993: 167). Poehner (in progress) is currently conducting an interactionist DA study of oral proficiency among advanced L2 learners of French. We strongly encourage other SLA and LT researchers to vigorously explore the full potential of DA.

Endnotes

¹Please cite as: Poehner, M. E. & Lantolf, J. P. (2003). *Dynamic Assessment of L2 Development: Bringing the Past into the Future*. (CALPER Working Paper Series, No. 1). The Pennsylvania State University: Center for Advanced Language Proficiency, Education and Research.

²As we will discuss shortly, there is yet a third way, and more idiosyncratic, way in which feedback can be offered in DA. This difference characterizes what we will refer to as the interventionist from the interactionist approach to DA.

³Lidz and Gindis (2003: 105) point to a similar distinction within DA that developed in Russia. One foreground assessment of learning ability and the other, more intimately connected to Vygotsky's theory, stresses teaching and learning in the ZPD.

⁴From Vygotsky's perspective the way in which the child interprets what he or she is seeing when alone as in the TV example is very much mediated by what has been internalized from previous mediated interactions with others. Space is because for Vygotsky, history is a critical feature of development and for this reason he argues that our activity is always and everywhere mediated.

⁵For a full discussion of imitation and internalization see Vygotsky (1987, 1998), Tomasello (1999), and as it relates to language learning Lantolf (2003).

⁶Aljaafreh and Lantolf (1994) and Swain and Lapkin (2002), while not directly situated within the DA literature, do focus on mediation in the ZPD and the psychological processes that are central to the interactionist orientation. Space does not permit us to detail these studies here, but we suggest that they be consulted by those interested in pursuing L2 DA research.

⁷ Lidz (2003) notes that in the former Soviet Union, standardized testing was proscribed in favor of dynamic approaches to assessment from the 1930s to the early 1980s and DA as the "only paradigm accepted in psychology and remedial education" (p. 105).

⁸ Interestingly, Alan Davies, in reacting to an earlier version of the present article presented at BAAL 2003, stated precisely that DA is teaching not testing.

References

- Alderson, J.C. and D.Wall (1993). Does washback exist? *Applied Linguistics* 14: 115-29.
- Antón, M. (2003). Dynamic assessment of advanced foreign language learners. Paper presented at the American Association of Applied Linguistics, Washington, D.C., March, 2003.
- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman, L. F. and A. D. Cohen (1998). *Interfaces Between Second Language Acquisition and Language Testing Research*. Cambridge: Cambridge University Press.
- Baker, L. and A. Brown. (1984). Metacognitive skills of reading. In *Handbook on Research in Reading*. D. P. Pearson (Ed.). New York: Longman.
- Brindley, G. (1998). Describing language development? Rating scales and SLA. In *Interfaces between Second Language Acquisition and Language Testing Research*. L. F. Bachman and A. D. Cohen (Eds.). Cambridge: Cambridge University Press.
- Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist* 32: 513-531.
- Brown, A. and R. A. Ferrara (1985). Diagnosing zones of proximal development. "In *Culture, Communication and Cognition. Vygotskian Perspectives*. J. V. Wertsch (Ed.). Cambridge: Cambridge University Press.
- Büchel, F.P. and U. Scharnhorst (1993). The Learning Potential Assessment Device (LPAD): Discussion of theoretical and methodological problems. In *Learning Potential Assessment: Theoretical, Methodological and Practical Issues*. J.H.M Hamers, K. Sijtsma and A.J.J.M. Ruijssemaars (Eds.). Amsterdam: Swets & Zeitlinger.
- Budoff, M. (1968). Learning potential as a supplementary testing procedure. In *Learning Disorders. Vol. 3*. J. Hellmuth (Ed.). Seattle, WA: Special Child.
- Budoff, M. (1987a). The validity of learning potential assessment. In *Dynamic Assessment: An Interactive Approach to Evaluating Learning Potential*. C. S. Lidz (Ed.). New York: The Guilford Press.
- Budoff, M. (1987b). Measures for assessing learning potential. In *Dynamic Assessment: An Interactive Approach to Evaluating Learning Potential*. C. S. Lidz (Ed.). New York: The Guilford Press.
- Budoff, M. and M. Friedman (1964). "Learning potential" as an assessment approach to the adolescent mentally retarded. *Journal of Consulting Psychology* 28: 434-439.
- Campione, J. C., A. L. Brown, R. A. Ferrara, and N. R. Bryant (1984). The Zone of Proximal Development: Implications for Individual Differences and Learning. In *Children's Learning in the "Zone of Proximal Development"*. B. Rogoff & J. V. Wertsch (Eds.). San Francisco: Jossey-Bass.
- Candlin, C. N. (2001). Afterward: Taking the curriculum to task. In *Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing*. M Bygate, P. Skehan and M. Swain (Eds.). London: Longman.
- Carlson, J. and K.H. Wiedl (2000). The validity of dynamic assessment. In *Dynamic Assessment: Prevailing Models and Applications*. C.S. Lidz and J.G. Elliott (Eds.). Amsterdam: Elsevier.
- Chalhoub-Deville, M. (2001). Task-based assessments: Characteristics and validity evidence. In *Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing*. M Bygate, P. Skehan and M. Swain (Eds.). London: Longman.
- Corman, L. and M. Budoff (1973). *A comparison of group and individual training procedures on the Raven Learning Potential Measure*. RIEPrint # 56 Cambridge, MA: Research Institute for Educational Problems.
- Elkonin, D. (1998). Epilogue. *The Collected Works of L. S. Vygotsky. Vol. 5. Child Psychology*. R. W. Rieber (Ed.) with a prologue by C. Ratner. New York: Plenum.
- Feuerstein, R., Y. Rand, and J. E. Rynders (1988). *Don't Accept Me as I Am. Helping Retarded Performers Excel*. New York: Plenum.
- Feuerstein, R., Y. Rand, M. B. Hoffman, and R. Miller (1980). *Instrumental Enrichment*. Baltimore, MD: University Park Press.
- Feuerstein, R., R. Miller, M. B. Hoffman, Y. Rand, Y. Mintzker, and M. R. Jensen (1981). Cognitive modifiability in adolescence: Cognitive structure and the effects of intervention. *The Journal of Special Education* 15: 269-287.
- Feuerstein, R., Y. Rand, and M. B. Hoffman (1979). *The Dynamic Assessment of Retarded Performers: The Learning Potential Assessment Device, Theory, Instruments, and Techniques*. Baltimore: University Park Press.
- Fredriksen, J. and A. Collins (1989) A system approach to educational testing. *Educational Researcher* 18: 27-32.
- Glutting, J.J. and P.A. McDermott. (1990). Principles and problems in learning potential. In *Handbook of Psychological and Educational Assessment of Children. Intelligence and Achievement*. C.R. Reynolds and R.W. Kamphaus. New York: Guilford.
- Guthke, J. (1982). The Learning Test Concept—An alternative to the traditional static intelligence test. *The German Journal of Psychology* 6: 306-324.
- Guthke, J., A. Heinrich, and M. Caruso (1986). The diagnostic program of "syntactical rule and vocabulary acquisition" - A contribution to the psychodiagnosis of foreign language learning ability. In *Human Memory and Cognitive Capabilities. Mechanisms and Performances*. F. Klix and H. Hagedorf (Eds.). Amsterdam: Elsevier.
- Hanson, F. A. (1993). *Testing Testing. Social Consequences of the Examined Life*. Berkeley, CA: University of California Press.
- Haywood, H. C., A. L. Brown and S. Wingenfeld (1990). Dynamic approaches to psychoeducational assessment. *School Psychology Review* 19: 411-422.
- Karpov, Y. V. and B. Gindis (2000). Dynamic assessment of the level of internalization of elementary school children's problem-solving activity. In *Dynamic Assessment: Prevailing Models and Applications*. C. Lidz and J. G. Elliott (Eds.). Amsterdam: Elsevier.
- Kozulin, A. and E. Garb (2002). Dynamic assessment of EFL text comprehension of at-risk students. *School Psychology International* 23: 112-127.
- Krashen, S. D. (1981). *Second Language Acquisition and Second Language Learning*. Oxford: Pergamon Press.
- Krashen, S. D. (1983). *Principles and Practice in Second Language Acquisition*. Oxford: Pergamon Press.
- Kvale, S. (1994). Ten standard objections to qualitative research interviews. *Journal of Phenomenological Psychology* 25: 147-173.

- Lantolf, J. P. (2003). Intrapersonal communication and internalization in the second language classroom. In *Vygotsky's Theory of Education in Cultural Context*. A. Kozulin, V. S. Ageev, S. Miller, & B. Gindis, (Eds.) Cambridge: Cambridge University Press.
- Lantolf, J. P. and W. Frawley (1985). Oral proficiency testing. A critical analysis. *The Modern Language Journal* 69: 337-345.
- Lantolf, J. P. and W. Frawley (1988). Proficiency: Understanding the construct. *Studies in Second Language Acquisition* 10: 181-195.
- Leont'ev, D. A. (2002). Activity theory approach: Vygotsky in the present. In *Voices within Vygotsky's Non-classical Psychology: Past, Present, Future*. D. Robbins and A. Stetsenko (Eds.). New York: Nova Science.
- Lidz, C. S. (1991). *Practitioner's Guide to Dynamic Assessment*. New York: Guilford.
- Lidz, C. S. and B. Gindis (2003). Dynamic assessment of the evolving cognitive functions in children. In *Vygotsky's Educational Theory in Cultural Context*. A. Kozulin, B. Gindis, V. S. Ageyev, S. M. Miller (Eds.). Cambridge: Cambridge University Press.
- Lumley, T. and A. Brown (1996). Specific-purpose language performance tests: Task and interaction. *Australian Review of Applied Linguistics* 13: 105-136.
- Lunt, I. (1993). The practice of assessment. In *Charting the Agenda. Educational Activity after Vygotsky*. H. Daniels (Ed.). London: Routledge.
- Luria, A. R. (1973). *The Working Brain*. New York: Basic Books.
- Luria, A. R. 1976. *Cognitive Development. Its Cultural and Social Foundations*. Cambridge, MA: Harvard.
- McNamara, T. (2001). Language assessment as social practice: Challenges for research. *Language Testing* 18: 334-349.
- McNamara, T. (1997). 'Interaction' in second language performance assessment: Whose performance? *Applied Linguistics* 18: 446-466.
- Messick, S. (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. In *Test Validity*. H. Wainer and H. I. Braun (Eds.). Hillsdale, NJ: Erlbaum.
- Messick, S. (1996) Validity and washback in language testing. *Language Testing* 13: 241-257.
- Minick, N. (1987). Implications of Vygotsky's theories for dynamic assessment. In *Dynamic Assessment: An Interactive Approach to Evaluating Learning Potential*. C. S. Lidz (Ed.) New York: The Guilford Press.
- Moss, P. A. (1994). Can there be validity without reliability? *Educational Researcher* 23: 5-12.
- Newman, F. and L. Holzman (1993). *Leo Vygotsky, Revolutionary Scientist*. London: Routledge.
- Pena, E. D. and R. B. Gillam (2000). Dynamic assessment of children referred for speech and language evaluations. In *Dynamic Assessment: Prevailing Models and Applications*. C. Lidz and J. G. Elliott (Eds.). Amsterdam: Elsevier.
- Pienemann, M. (1998). *Language Processing and Second Language Development. Processability Theory*. Amsterdam: John Benjamins.
- Poehner, M. (in progress). *Dynamic Assessment of Advanced L2 Learners of French*. Ph.D. dissertation. Penn State University.
- Polkinghorne, D. E. (2003). Generalization and qualitative research: Issues of external validity. In *Validation of Knowledge Claims in Human Science*. J. Lindén and P. Szybek (Eds.). Lyon: L'interdisciplinaire.
- Ryan, J. and Lafford, B. (1992). Acquisition of lexical meaning in a study abroad environment: SER and ESTAR and the Granada experience. *Hispania* 75: 714-722.
- Schneider, E. (1999). *Multisensory Structured Metacognitive Instruction: An Approach to Teaching German as a Foreign Language to At-risk Students*. Frankfurt am Main: Peter Lang.
- Schneider, E. and L. Ganschow. (2000). Dynamic assessment and instructional strategies for learners who struggle to learn a foreign language. *Dyslexia* 6: 72-82.
- Scribner, S. and M. Cole. 1981. *The Psychological Consequences of Literacy*. Cambridge, MA: Harvard.
- Shohamy, E. (1988). Closing the gap between research and practice: Applications of research results to classroom testing. Revised version of a paper delivered at the English Teachers' Association of Israel International Conference, Jerusalem.
- Shohamy, E. (2001). *The Power of Tests*. Mahwah, NJ: Erlbaum.
- Shehan, P. (2001). Tasks and language performance assessment. In *Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing*. M Bygate, P. Skehan & M. Swain (Eds.). London: Longman.
- Snow, R.E. (1990). Progress and propaganda in learning assessment. *Contemporary Psychology* 35: 1134-1136.
- Sparks, R. L. and L. Ganschow (1993a). Searching for the cognitive locus of foreign language learning difficulties: Linking first and second language learning. *The Modern Language Journal* 77: 289-302.
- Sparks, R. L. and L. Ganschow (1993b). The impact of native language learning problems on foreign language learning: Case study illustrations of the linguistic coding deficit hypothesis. *The Modern Language Journal* 77: 58-74.
- Spence-Brown, R. (2003). *Authentic Assessment? The Implementation of an 'Authentic' Teaching and Assessment Task*. Unpublished doctoral dissertation. University of Melbourne.
- Sternberg, R.J. (2000). Prologue to dynamic assessment: Prevailing models and applications. In *Dynamic Assessment: Prevailing Models and Applications*. C. Lidz and J. G. Elliott (Eds.). Amsterdam: Elsevier.
- Swain, M. (1984). Large-scale communicative language testing: Case study. In *Initiatives in Communicative Language Teaching. A Book of Readings*. S.J. Savignon and M.S. Berns (Eds.). Reading, MA: Addison-Wesley.
- Swain, M. (1993). Second language testing and second language acquisition: Is there a conflict with traditional psychometrics? *Language Testing* 10: 193-207.
- Swain, M. (2001). Examining dialogue: Another approach to content specification and to validating inferences drawn from test scores. *Language Testing* 18: 275-302.
- Swain, M. and S. Lapkin (2000). Focus on form through collaborative dialogue: Exploring tasks effects. In *Researching Pedagogic Tasks: Second Language Learning, Teaching, and Testing*. M. Bygate, P. Skehan, and M. Swain (Eds.). London: Longman.
- Swain, M. and S. Lapkin (2002). Talking it through: Two French immersion learners' response to reformulation. *International Journal of Educational Research* 37: 285-304.
- Tomasello, M. (1999). *The Cultural Origins of Cognition*. Oxford: Oxford University Press.
- Tulviste, P. 1991. *The Cultural-Historical Development of Verbal Thinking*. Commack, NY: Nova Science Publishers.
- Tzuriet, D. (2001). *Dynamic Assessment of Young Children*. New York: Kluwer/Plenum.
- Valsiner, J., 2001. Process structure of semiotic mediation in human development. *Human Development* 44: 84-97.
- Van der Veer, R. and J. Valsiner (1991). *Understanding Vygotsky*. Oxford: Blackwell.
- Vygotsky, L. S. (1956). *Isbrannye psikhologicheskie issledovaniya* [Selected psychological investigations]. Moscow: Izdatel'stvo Akademii Pedagogicheskikh Nauk SSSR.
- Vygotsky, L. S. (1986). *Thought and Language*. Newly revised and edited by A. Kozulin. Cambridge, MA: MIT Press.
- Vygotsky, L. S. (1997). The crisis in psychology. In *The Collected Works of L. S. Vygotsky*. Vol. 3. *Problems of the theory and history of psychology*. Including the chapter on the crisis in psychology.. R. W. Rieber (Ed.). New York: Plenum.