

**Center for Advanced Language Proficiency Education and Research**

a National Language Resource Center

**Dynamic Assessment and the  
Problem of Validity in the L2 Classroom**

by  
**Matthew E. Poehner**  
The Pennsylvania State University

No. 10, May 2008

PENNSSTATE



**CALPER Working Paper Series**

# Dynamic Assessment and the Problem of Validity in the L2 Classroom<sup>1</sup>

Matthew E. Poehner  
The Pennsylvania State University

## Abstract

To date, assessment validation has primarily been explored within the context of large-scale standardized testing and little attention has been given to the validity of classroom-based assessment. The assumption that processes of validating standardized tests can be unproblematically applied to classroom assessments has begun to be questioned on the grounds that these two types of assessment differ in fundamental ways (Moss 2003). The former seeks to isolate abilities, which are conceived as relatively stable, discrete traits that can be sampled and measured. The latter, in contrast, is less interested in observed consistencies in performance than in helping individuals improve their functioning, and this entails understanding the processes of development. These divergent goals have important implications for how assessment validation is approached. In the testing literature, Messick's (1989) unified validity model remains highly influential. For Messick test validation is a scientific enterprise employing methods akin to hypothesis generation and evaluation in order to establish the legitimacy of interpretations and uses of test scores, which are taken to be samples of individuals' abilities. In the classroom, assessment need not be conceptualized as a formal, one-time standardized procedure but is often integrated with – and in support of – teaching and learning. Reorienting assessment from a measurement activity to one focused on learner development implies a shift from mere observation of performance to cooperation with learners and intervention. A theoretical framework for development-oriented assessment is articulated in Vygotsky's (1987) Sociocultural Theory and in the writings of his colleague, A. R. Luria (1979), on "romantic science." Their approach, known as Dynamic Assessment (DA), seeks not to describe performance consistencies but to engage in dialogue with learners in order to arrive at a diagnosis of the sources of performance problems and a prognosis for their remediation based on learner responsiveness during the interaction. In this way, teaching and assessment are seamlessly integrated. Drawing on examples of DA with L2 learners, it is argued that the validity of this activity is ultimately derived from its success in promoting learner development.

Language assessments are routinely carried out by professionally trained testers and psychometricians as well as by classroom language teachers with little background in assessment principles.<sup>2</sup> In both cases, assessment is a motivated activity – that is, assessments are conducted to gain insights into individuals' knowledge and abilities for the purpose of making various kinds of decisions ranging from screening and selecting applicants to a program of study (as in the case of many large-scale standardized tests such as the TOEFL) to determining whether additional instructional time should be devoted to a given topic

<sup>1</sup> This publication was funded by a grant from the United States Department of Education (CFDA 84.229 P229 A060003). However, the contents do not necessarily represent the policy of the Department of Education, and one should not assume endorsement by the Federal Government.

<sup>2</sup> Please cite as: Poehner, M. E. (2008). *Dynamic Assessment and the Problem of Validity in the L2 Classroom*. (CALPER Working Paper Series, No. 10). The Pennsylvania State University: Center for Advanced Language Proficiency Education and Research

Throughout this paper I use the term testing to refer to the use of formal assessment instruments and procedures on a large-scale, as in the case of the TOEFL or IELTS. Assessment is a broader term that includes tests as well as other approaches to evaluating learning such as portfolios, projects, role plays, presentations, etc. Classroom-based assessments I define as assessments that are designed and implemented (and whose outcomes are interpreted) by classroom teachers rather than testing professionals.

(as often occurs in classroom settings). Assessment purposes have traditionally been categorized as either *summative*, that is, indicative of overall achievement and past learning, or *formative*, in which case the assessment is intended to inform teachers and students about learners' progress and problems (Torrance & Pryor, 1998). Both of these purposes posit very different relationships between assessment and teaching and learning. In the case of summative assessments, assessment is a standalone activity that reports the products of learning but is not intended to impact teaching and learning directly, and when impact occurs as in the case of washback (Alderson & Wall, 1993; Cheng 2005; Cheng, Watanabe & Curtis 2004), it is often framed in a negative light, with fears that education will be reduced to test preparation. Formative assessments, in contrast, are explicitly connected to teaching and learning. At a minimum, formative assessments are understood to be part of an instructional cycle whereby assessment results inform subsequent teaching, but they may also co-occur with teaching as when teachers offer feedback during informal classroom activities and both teachers and learners reflect on performance (see Ellis, 2003).

All assessments, regardless of their purpose, must address the matter of validity as this concerns the appropriateness of decisions based on assessment information. Bachman (2000) has characterized validation as a process of deciding whether what a test measures is really worth counting. In other words, validity cuts to the heart of the assessment enterprise by challenging assessors to defend their interpretations of assessment performance and to consider the consequences of assessment for individuals, programs and institutions, and society. Despite important differences between large-scale testing and classroom-based formative assessments, discussions of assessment validity (in applied linguistics, as well as in education and psychology more generally) have primarily centered on the former, tacitly assuming that whatever models are worked out for standardized tests will also be suitable to the classroom. This bias might be defended on the grounds that life-altering decisions are made on the basis of "high-stakes" tests. It is also true that the concept of validity, currently associated with issues of fairness and ethics (McNamara 2001; Lynch 2001; Davies 2003), may be more amenable now to classroom applications than in its inception as an essentially statistical property of tests. Nevertheless, it remains the case that comparatively little attention has been paid to validity with regard to classroom assessment. Indeed, there has been growing recognition among L2 scholars in recent years of the high stakes that classroom assessments carry for teachers and learners (Rea-Dickins 2001; Rea-Dickins & Gardner 2000) but this is a highly under-researched area.<sup>3</sup>

Teadale and Leung (2000), writing in the context of national curricular reform in the UK that included both large-scale testing as well as formative assessments, raised the issue of whether classroom assessments could or even should adhere to psychometric standards of validity. These authors point out that in the classroom, where assessment is intended "to meet the needs of pedagogy," psychometric notions of validity may need to be reconceptualized; they also urge practitioners to articulate theoretical and ideological connections between psychometric and formative understandings of validity (Teadale & Leung 2000: 165). Others, such as Torrance (1995), warn against evaluating new assessment approaches according to psychometric criteria, particularly when they are intended to serve a pedagogical function as this contrasts sharply with the measure-

<sup>3</sup> See, however, the special issue of *Language Testing* (2004) on classroom-based assessment.

ment focus of standardized testing. Along similar lines, Moss (2003) has voiced doubt that approaches to validating standardized tests are appropriate to formative assessments, and has expressed concern that importing a psychometric model into the classroom would also entail an "artificial" separation of what she sees as the complementary processes of assessment and teaching (p. 15). She summarizes the issue by asking:

To what extent does our understanding of validity in the measurement profession assure that the relevant issues are addressed in classroom assessment and what role might other theoretical perspectives play in providing a more robust validity framework to guide thinking and action? (Moss 2003, p. 15)

In this paper, I offer a response to Moss's query by approaching classroom assessment and its validity from a Vygotskian theoretical perspective. A substantial body of work now exists that employs Vygotsky's (1987) Sociocultural Theory of Mind to illuminate teaching practices and their impact on learner development (Daniels 1993; Karpov 2005; Kozulin, Gindis, Ageyev & Miller 2003; Lantolf & Poehner 2008; Wells & Claxton 2002). In addition, the research of Vygotsky and his close collaborator, A. R. Luria, has led to a conceptualization of teaching and assessment in which these form not complementary processes as Moss suggests but rather a dialectical unity that is focused on learner development (Luria 1961; Vygotsky 1998). In this framework, known as Dynamic Assessment (DA), the assessor, or mediator, dialogically engages with learners to formulate a diagnosis that includes fully formed abilities as well as those that are still developing (Sternberg & Grigorenko 2002; Haywood & Lidz 2007), or as Vygotsky (1978) put it, are in the process of "ripening." More important, DA interventions simultaneously support development by providing assistance that is sensitive to learners' emergent needs. Rather than presenting a stable portrait of learners' abilities, DA actively seeks to promote change through mediator-learner cooperation.

The dialectical integration of teaching and assessment that DA represents offers a powerful framework for the classroom, a matter that has begun to be explored by L2 researchers (Ableeva 2008; Lantolf & Poehner 2004; Leung 2007; Poehner 2007, 2008). Reframing assessment as a development- rather than measurement-oriented activity is clearly in line with the formative goals of the classroom setting but it also lends considerable weight to the aforementioned concerns over psychometric criteria for validity. As Lantolf and Poehner (2007) propose, validating the activity of teaching-assessment requires interpreting its impact on learner development, and just as teaching-assessment in the classroom is ongoing so too must be its validation. Validity, then, can no longer be a concern exclusive to testing specialists but is fundamental to all classroom practitioners.

To be sure, the proposed dialectical view of validity is a clear departure from current thinking in psychometrics, where testing and test validation are likened to scientific processes akin to investigation in physics, chemistry, or biology (Messick 1988, 1989, 1995). At issue is not whether assessment can be scientific but the model (or models) of science most appropriate to understanding human abilities. In this regard, Vygotsky (1997) cautioned psychology strongly against the tendency to "borrow" principles, constructs, laws, and hypotheses that have borne success in the natural sciences:

The theses and constructions of these highly developed, firmly grounded sciences are, of course, methodologically elaborated in an infinitely more precise way than the theses of a psychological school which by means of newly created and not yet systematized concepts is developing completely new areas...[and yet] as the borrowing here comes from other sciences, the material turns out to be more foreign, methodologically heterogeneous, and the conditions for appropriating it become more difficult (p. 269).

More strongly still, he submits that when otherwise sound methods are applied to an object of study to which they do not correspond, the result is an "illusion of science" that can be salvaged neither by "the mathematical formulas nor the precision equipment" of other disciplines (Vygotsky 1997, pp. 269-270).

Since Vygotsky's time, challenges to the adoption of natural science research methods in the humanities have continued, with particular concern expressed for the ways in which researchers are compelled to construct the object of study (i.e., the human mind in the case of psychology and education) in order for it to be amenable to controlled experimentation and measurement (Danziger 1990; Ratner 2002, 2006; see also Lantolf 2000 for a critique of experimental methods in SLA research). In the realm of testing, the quest for rigorous scientific methods has resulted in a preoccupation with standardized procedures, testing individuals in isolation, and statistical manipulation of scores. These practices are often deemed necessary due to the large-scale implementation of many standardized tests, although the widespread interest in the impact of tests and the need for professional testing ethics suggests a growing awareness of their potentially harmful consequences. My purpose here is neither to attack nor defend standardized testing and psychometric understandings of validity but rather to bring to light underlying assumptions in this paradigm that are not conducive to the goal in classroom assessment of supporting learner development.

In what follows, I begin with a brief overview of the origins of psychometric testing and validity. With this discussion as background, I offer a more concrete critique of the application of psychometric models to address the validity of classroom-based assessment. I then turn to DA as a theoretically-driven approach to classroom teaching and assessment that also provides the kind of "robust validity framework to guide thinking and action" that Moss (2003, p. 15) has called for.

### Assessment as Science

Several useful overviews of validation theory are available both in the general educational and psychological research literatures (Anastasi & Urbina 1996; Cronbach 1990) as well as in language testing (McNamara & Roever 2006), and the reader is referred to these for a more comprehensive treatment of this topic than can be offered in the present paper. While the concept of validity has changed substantially over the years, there remains an underlying acceptance of psychoeducational testing as a scientific enterprise and validation as a form of hypothesis verification, a metaphor made explicit by Messick (1989). Framing assessment in natural science terms necessarily puts it at odds with teaching: while the former seeks to obtain uncontaminated samples of ability that can be represented quantitatively, the latter is interested in change, that is, in supporting the development of learner abilities. While a measurement orientation to assessment may be justifiable in certain contexts, as when dealing with large numbers

of individuals (although see Shohamy 2001, for counterarguments), it is inconmensurate with the goals of classroom practice. However, before proposing how assessment and teaching might be more fully aligned, it is instructive to review the means by which they were separated in the first place. To do so, we will turn briefly to the history of psychology, and in particular to the search for method within that discipline.

A good deal has been written regarding the appropriateness of importing methods from the natural sciences to the human sciences (see for instance Yngve, 1996; Danziger, 1997; Cole 1996), and the implications of this work for theory building, research methodologies, and practice in applied linguistics has been debated (Long 1993; Gregg 1993; Rampton 1995; Block 1996; Lantolf 1996). It is not my intention to attempt to disentangle the complex issues at stake in this debate. My interest rather is in the implications of framing classroom assessment practices, and the validation of those practices, according to a model of science that calls for the isolation and quantification of discrete variables. The central question is, What assumptions about mental abilities must be made for psycho-educational assessment to be a science of measurement (as Messick contends)? Exploring this question will better position us to consider the relevance of these assumptions for the classroom.

### The Need to Measure

Polkinghorne (1983, p. 15) observes that the debate around methodology in the human sciences is centuries old, arising not long after the emergence of empirical research in Europe during the Middle Ages and Renaissance. He explains that it was Galileo who was primarily responsible for introducing the European scientific community to the notion that the natural world is not random but obeys consistent patterns or laws. Galileo insisted that through appropriate methods, it was possible to observe, describe, and predict natural phenomena, and that mathematics offered the most efficient and eloquent means to represent their regularity (Polkinghorne 1983, p. 16). Thomas Hobbes, a contemporary and colleague of Galileo, first raised the question of whether human behavior might be subject to these same laws and thus amenable to the same means of inquiry.<sup>4</sup>

Nonetheless, it was not until the Nineteenth Century that researchers began to think seriously about the possibility of empirically studying the human mind. By that time the methodology worked out by Descartes had produced impressive results in the natural sciences. Sullivan (1984, p. 21) explains that on an epistemological level, Cartesian thinking was largely responsible for the rise of both rationalism and empiricism – that is, methods of inquiry based upon observation, logical analysis, and causal explanation. Together rationalism and empiricism "provide an interpretation of what it means to account for something, both in the sense of telling what a phenomenon or event is like, which is description, and in the sense of establishing why it had to follow from something else, which is explanation in the strictest sense" (ibid.). Descartes, however, was not satisfied with a mode of inquiry that produced conclusions that seemed ambiguous or that might be disputed. In his *Discourse on Method*, he outlined a series of steps that must be followed in order to "elimi-

<sup>4</sup> In fact, Hobbes went so far as to reject the classic Cartesian mind/body split, which held that the mind and body are composed of different substances with the former being mental and the latter physical. According to Descartes, this meant that only the body, being made of physical substance, could be studied using the rigorous methods of science, but Hobbes countered by suggesting that the mind, too, is part of nature and so is no different from any other natural phenomena (Polkinghorne, 1983, p. 17).

nate interpretation" and assure absolute certainty (Sullivan, 1984, p. 20). For Descartes, as long as interpretation was part of the scientific process one could never arrive at the truth. His solution, the now famous Cartesian scientific method, included the generation of hypotheses to be proved or disproved, the atomistic reduction of problems or phenomena to their most basic elements, analysis of elements from the most basic to most advanced, and a reliance on the certitude of measurement and mathematical expression (ibid.).

It is important to note, as does Gadamer (1975), that this dominant way of doing science essentially forms research questions in such a way that they may be efficiently investigated using mathematical means, including measurement, and that while this is certainly not the only approach to empirical systematic inquiry, it is effective for the study of natural phenomena. Of course, Descartes' was not the only voice concerned with methodological issues. Indeed, the German Romantic tradition associated with Herder and others rejected the universalism of Descartes' approach particularly as applied to the study of human phenomena (Berlin 2000). Yet it was the Cartesian scientific method that produced such astounding technological innovations and advances in the study of the natural world. According to Cole (1996), the allure of these methods for psychologists wishing to elevate their discipline to the level of a "general" science that could discover "universal mechanisms" of the human mind was too great to ignore (p. 32).

Cole traces the first serious work on developing a scientific methodology in psychology to the work of Wilhelm Wundt at the end of the Nineteenth Century. Wundt was concerned with establishing psychology as a separate discipline from philosophy, which also attempted to understand the mind. Viewed as the father of experimental psychology, Wundt established a lab for conducting research that followed the scientific practices of controlling conditions, isolating discrete variables, and quantifying and measuring phenomena.<sup>5</sup> In Wundt's day, however, the range of objects of psychological inquiry was still relatively limited, not yet including the many classic psychological constructs such as intelligence, motivation, and attention. Danziger (1990, 1997) explains that in North America in the early Twentieth Century behaviorist psychologists became interested in the burgeoning field of statistics and particularly in the notion of variables. Given what Polkinghorne (1983, p. 23) has described as American psychology's "easy embrace of the scientific and technological perspective," it was not long before the classic stimulus-response model was re-imagined as a relation of independent to dependent variables (Danziger 1997, p. 164). An important consequence of this reconceptualization is that it became possible to express individuals' responses to stimuli in mathematical terms, with numerical values being assigned to the quality of individuals' responses, and soon the abilities thought to underlie observable behavior, such as intelligence, motivation, memory, and attention, also came to be seen as variables in an elegant mathematical equation (see Danziger, 1990; Lantolf 2000). Moreover, just as variables in statistics may be substituted, controlled, and manipulated, the new psychological variables were similarly positioned to be amenable to purposeful manipulation and experimentation in much the

<sup>5</sup> Accepting the Cartesian mind-body split, Wundt proposed two types of psychology, folk psychology and physiological psychology. Folk psychology was concerned with the study of language, culture, and myths, which Wundt proposed could provide insights into higher mental processes, while physiological psychology dealt with immediate experiences of the world through the senses. For Wundt, only the latter could be a full scientific discipline because physiological phenomena have a biological basis and are therefore amenable to study through scientific means (Polkinghorne, 1983, pp. 34-35).

way that temperature, light, pressure, and access to water and minerals are all variables that can be purposefully altered in natural science research. The era of widespread standardized testing – that is, measuring the amounts of abilities individuals possess – began.

According to Gould (1996), the first application of standardized testing of abilities occurred in the context of immigration and later as a basis for determining work assignments within the U. S. military. In both cases testing was tied heavily to the eugenics movement, as it provided a pseudoscientific justification for racial prejudices and social inequities (see also Hanson 1993). As Sacks (1999) explains, standardized testing offers an appearance of objectivity and fairness because all examinees are subjected to precisely the same testing instruments, conditions, and procedures and all test results are interpreted following the same methods. The underlying assumption is that the individual has no effect on the conditions – all people will respond in similar ways regardless of social class, education level, linguistic, ethnic or religious background, age group, gender, or sexual orientation. Testing proponents thus allege its inherently unbiased, meritocratic nature, despite considerable evidence that tests often advantage individuals from privileged social backgrounds and thereby perpetuate class differences (Sacks, 1999, p. 15). Nonetheless, since the end of the Second World War, standardized testing for educational purposes has become the accepted and preferred method to accept or reject applicants, to mark academic achievement, to certify competency, to place learners or promote them within a program of study, and to disperse merit-based awards. To be sure, the methods for administration and interpretation associated with this mode of assessment have removed it completely from the goals of teaching and learning. Isolating individuals from their peers, denying access to cultural forms of support, imposing time limits, ignoring differences in learning styles, dismissing partial correctness or understanding, and refusing requests for help are all antithetical to good teaching but definitive of standardized testing. I should stress again that this is not intended as a condemnation of psychometrics, but rather I wish to underscore that teaching and testing have become separate endeavors. Testing has emerged as its own enterprise with extensive commercial applications and considerable power over individuals' lives. Just as the natural sciences have been central to the formulation of abilities in psychoeducational measurement, they have been equally influential in how the matter of testing validity has been approached.

#### Validity: From Correlations to Consequences

Contemporary discussions of validity can be traced back to the influential work of Cronbach and Meehl (1955), who proposed that testers needed to move beyond the view of validity as a statistically derived property of tests. Up to that point, the numerical scores produced by large-scale tests had led to correlational studies of validity. This understanding of validity, generally referred to as *criterion validity*, was determined by correlating the results of one test with those of another test, either given at the same point in time (*concurrent validity*) or at some later date (*predictive validity*). However, the correlational approach had begun to raise concerns among test developers and test users, as one could never be sure that any test was really valid since its validity depended upon the validity of other measures whose validity in turn depended upon those of others, and so on. Testers needed a less arbitrary method to establish that their procedures

were indeed measuring what they were intended to and that their measurements were relatively free from error introduced by other variables. The alternative put forth by Cronbach and Meehl was to focus on *construct validity*, which involves accurately and appropriately theorizing the "trait or quality" believed to underlie test performance (Cronbach and Meehl, 1955, p. 283). The importance of this proposal was that it broadened the scope of testing beyond scores and towards the meaning that can be attributed to those scores. What is of interest is not an individual's performance on a given test but what this might reveal about his/her abilities and the predictions that can be made regarding his/her likely performance in other settings. Reconciling validity along these lines meant that no test in itself is valid, rather it may be used and interpreted in ways that are more or less valid. This reorientation of validation studies set the stage for Messick's (1989) highly influential validity framework, which underscores the need for assessors to build arguments to support the use and interpretation of assessments.

Messick understood validity as a unitary concept such that the various types of validity, including content validity, criterion validity, and face validity, are facets of, or are encompassed by, construct validity. From this perspective, construct validity forms the basis of the entire assessment enterprise. It includes carefully articulating a theory of the abilities, or construct, to be assessed, the evidence to be used to make claims about test-takers, and consideration of the potential consequences of the assessment for individuals, institutions, and society. Messick's multi-faceted validity model is reproduced in Table 1.

TEST INTERPRETATION	TEST USE
EVIDENTIAL BASIS	Construct validity
CONSEQUENTIAL BASIS	Value implications Social consequences

Table 1: Facets of validity (Messick, 1989: 20)

In this model, validation is a process of grounding specific interpretations and uses of assessments in evidence obtained through the procedures. This process orientation to validity is also characteristic of Mislevy's (1996) and Kane's (1992, 2001) more recent argument-based approaches to validation. For these authors, validity involves a series of inferences in order to generalize instances of observed behaviors under specified testing conditions to a *criterion domain*, that is, to non-testing target situations. Typically, this involves modeling the demands of the target, including the knowledge and abilities thought to be required for successful performance, and then predicting how likely individuals are to meet these demands based on their assessment performance. As McNamara and Roever (2006, p. 16) explain, testing from this perspective represents a means of obtaining the necessary evidence to justify the predictions one wishes to make. Validation arguments are built by carefully and explicitly linking forms of evidence about individuals' abilities

to interpretations and predictions. Elsewhere, McNamara (2006, pp. 32-33) likens the process of validation in language assessment to a legal proceeding in which an accused individual's guilt or innocence is determined according to the arguments put forth by both the prosecution and defense attorneys, arguments that must take account of all available evidence. Similarly, the function of "empirical validation in language testing is to ensure the defensibility and fairness of interpretations based on test performance...the scrutiny of such procedures will involve both reasoning and examination of the facts" (ibid.).

While this process orientation to validity may represent an important advancement as far as standardized testing is concerned, the implications for the classroom are unclear. Validation in the psychometric paradigm continues to struggle with whether or not the characterization of individuals' abilities is appropriate while classroom activity is concerned with supporting development. In some sense, this is akin to two artists arguing over who has more accurately rendered an autumn landscape, with neither noticing that the seasons have changed. Vygotsky often invoked a medical metaphor relating traditional assessments to a diagnosis that merely restates the problem in scientific language but does not treat the problem (Vygotsky 1998, p. 205). We might extend this metaphor to assessment validity to say that psychometric validation is similar to a process of justifying medical diagnoses while teaching and learning are committed to overcoming the problem, and consequently undoing the diagnosis. As I explain below, Vygotsky's proposal was that together diagnosis and prognosis form a unity and that neither is complete without the other. However, more needs to be said regarding the relevance of psychometrically derived notions of validity to the classroom, and for this I return to the work of Moss (2003).

### Validity in the Classroom

Moss (2003, p. 14) outlines five key principles and practices in testing validation that set it apart from classroom-based assessment that is formative, or developmental, in nature. Foremost among these is her observation that assessment in the psychometric tradition is a discrete activity while in the classroom it is ongoing. As should be clear from the preceding discussion, the move to separate assessment from teaching was motivated by a desire to achieve scientific objectivity, but this has also had important consequences for how one attempts to understand learner development. Specifically, standardized testing usually involves a one-off observation of individuals performing tasks under assessment conditions (e.g., time constraints, denial of access to resources or materials, proscription of interaction with others). These observations are followed by a series of logical inferences that attempt to link performance to individuals' underlying abilities. Statistical procedures are called upon to support these inferences and to establish the degree of confidence one can have in generalizing claims about examinees' likely (future) performance in hypothetical contexts.

Gipps (1994), describing the advantages of classroom-based performance assessments, has suggested that generalizability is less of a concern in the classroom because assessment is not limited to single occurrences but can involve an accumulation of observations. Poehner (2007), proceeding from a Vygotskian perspective, argues that it is not whether generalizability is relevant to the classroom but rather how this construct is understood. He maintains that generalizing an interpretation of learners' abilities from a single performance makes little sense and that the focus of classroom interactions is

connecting each teaching-assessment episode to the next in order to continually work within learners' ZPDs. Within a Vygotskian framework, the classroom enables teachers to simultaneously promote learner development and track their progress over time. Claims about learner abilities are tied to concrete instances of classroom interactions and are construed as dynamic rather than stable. That is, learner development is profiled according to analysis of changes in performance over time, and this includes not only successful completion of specified tasks but also consideration of changes in the quality of support learners need, the degree of agency that characterizes their response to instructional support, and their ability to handle increasingly complex problems (Poehner, 2007).

Moss's other points regarding test validation further emphasize the discord between the psychometric paradigm and the classroom. She notes for instance that current concerns over the interpretation and use of information about individuals is limited to information obtained through formal tests, and that these insights into learner abilities are privileged over all other forms of knowledge, including teacher observations and records. A similar observation is made by Johnson, Rehn-Jordan, and Poehner (2005) in the context of university-level academic English language instruction in the U.S. These authors report that ESL teachers' end-of-term evaluations of their own learners were disregarded if they differed from the students' scores on an institutional-based TOEFL, and consequently students with poor grades but acceptable test scores were promoted to the next level of study but the reverse was not true. Sacks (1999, p. 14) found this same phenomenon with regard to the definition of gifted learners in many educational settings. He explains that giftedness is more often linked to high scores on standardized tests than to demonstrated excellence in the classroom, that is, to the "potential to achieve rather than achievement itself" (*ibid.*, italics in original). Following Sacks' analysis, the seemingly scientific objectivity of testing is given considerable credence, even when it conflicts with firsthand experience.

Moss continues that the object of testing – the construct to be measured – is an ability or knowledge believed to exist in an individual's head. While this model of cognition fits well with the traditional focus on solo performance in testing, it does not, as Moss points out, reflect the realities of the classroom, where collaboration and interaction are highly valued. Indeed, the intracranial model of cognition has been challenged from a number of perspectives that point towards the social origins of mental development (e.g., Lave & Wenger 1991; Salomon 1996; Wenger 1998). Research in this area is becoming increasingly influential in educational reform and pedagogical innovation, although as McNamara (2001) argues, its implications for language testing have not yet been addressed. Indeed, the assessment of language proficiency would certainly be approached very differently if it were understood not as a set of discrete traits but as a feature of the activities in which individuals engage, and therefore contingent upon context-specific affordances and negotiated among interactants (see van Lier, 2004). The matter of whether these assumptions are warranted and the restrictions they impose worthwhile in the context of standardized testing is a topic for another paper. The point here is that assessment that is part of the overall classroom activity of supporting learner development need not be bound to these assumptions and restrictions.

Moss additionally explains that testing and classroom practices differ in terms of how they approach assessment outcomes. The information about learners gleaned from tests is typically aggregated and summarized as scores, and sometimes convert-

ed into grades or percentile rankings. Consequences then come into play if it is found that the construct under assessment is not appropriately modeled or if other variables can be shown to affect assessment performance (Moss, 2003, p. 14). Here again, one could argue that although important insights into learner abilities and their development are no doubt lost when performance is reduced to a number or set of numbers, generating scores does offer a common metric for comparing individuals and this is the goal of large-scale testing. However, the classroom allows for multiple sources of information about learners, and more fine-grained analyses of learner abilities better position teachers to intervene in their development. This also has the effect of augmenting the importance of consequences, which gain their validity not simply from defensible interpretations but from the appropriateness of teaching aimed at promoting development.

As the preceding discussion makes clear, the formative or developmental orientation of classroom assessment does not mean that validity is irrelevant. Rather, reinterpreting teaching and assessment shifts validity from a discussion about an individual's performance to an activity of collaborating with the individual to help him/her perform better. This notion of a diagnostic intervention was a logical application of Vygotsky's theory and is central to DA. The next section describes the theoretical perspective on human abilities and their development that informs DA. This approach is then illustrated with examples of DA interactions involving L2 learners, and these examples serve as points of reference when our discussion returns to validity and what a development-oriented approach to validity might look like.

### Dynamic Assessment

Haywood and Lidz (2007) explain that DA is not a single method of assessment but refers to a wide range of practices that depart from traditional, or non-dynamic, assessments (NDA) by including intervention and learner responsiveness to intervention as essential to understanding learner abilities. Sternberg and Grigorenko (2002) note that DA interventions may be sandwiched between two NDAs, as in a pre-test – intervention – post-test model, or alternatively they may co-occur with the administration of the assessment itself, wherein a mediator interacts with the examinee whenever difficulties arise. Lantolf and Poehner (2004) further distinguish between what they term *interventionist* and *interactionist* DA, which concerns the quality of mediation offered to learners. In *interventionist* DA, practitioners devise scripted and standardized sets of prompts, hints, and leading questions, while proponents of *interactionist* DA advocate an open-ended, dialogic approach to mediation that allows mediators to respond flexibly to learners' changing needs and to pursue unforeseen problems. Outcomes of DA interactions may be reports, profiles, or sets of scores that typically include some or all of the following information: learner independent performance, degree of improvement with mediation, forms of mediation required, responsiveness to mediation, and ability to extend learning to more complex tasks (Poehner 2008). While these distinctions are helpful to understanding DA, it is important to bear in mind that DA is much more than a set of technical procedures – it is a fundamentally different way of thinking about assessment and its relation to teaching.



## Romantic Science

What sets DA apart from other conceptualizations of assessment is that it proceeds from a theoretical perspective on the human mind and its development that is grounded in a humanistic, or romantic, philosophy of science. Luria (1979) explains in his account of classical and romantic science that in the former scientists attempt to break down phenomena into their constituent parts, or their elemental units, and from this they abstract general theoretical principles or laws to explain the phenomena. Romantic scientists, in contrast, "want neither to split living reality into its elementary components nor to represent the wealth of life's concrete in abstract models that lose the property of the phenomena themselves" (Luria, 1979, p. 174). Good (2000) points out that Luria's most-celebrated studies are in fact in-depth portraits of individuals who were not simply subjects in the experimental sense but were implicated, along with Luria, in the research (p. 27). Good continues that this move from simple observation and description to interaction and collaboration was emblematic of Luria's commitment to exploring the "lifeworld" of individuals, an enterprise he perceived as at odds with "reductionist/mechanical devices... [including] computer simulation and mathematical models disconnected from the human condition" (ibid.).

Of course, the romantic view of science did not originate with Vygotsky and Luria but is associated Herder and others in the German Romantic tradition, as well as with the work of their Italian predecessor, Giambattista Vico. According to Berlin (2000), Vico objected to applications of Descartes' natural science methods to the study of humankind on the grounds that human beings differ from natural objects of study in an essential way, and this concerns causality. Vico contended that the natural world might have an underlying order organized according to causal relationships but human "purposive activity" must not be reduced to "mechanical cause and effect" and can only be interpreted in relation to intelligible motives and goals (Berlin 2000, p. 10). Descartes' reliance on observation, categorization, and measurement might be adequate to the task of explaining 'external' phenomena in the natural world but not to understanding the "effort, struggle, purposes, motives, hopes, fears, attitudes" that characterize human activity (p. 9). In contrast to the universalism of Descartes' methods, Vico advocated a new human science to obtain an 'inside' view of the worlds people inhabit:

The whole world of culture has, for certain, been produced by the physical and mental activity of man, and for this reason one can, and, in fact, has to, find its principles and regularities within the modes of existence of the spirit of the self-same people (Vico 1725, cited in Polkinghorne, 1983, pp. 20-21).

Herder and his colleagues were similarly concerned with the Cartesian enterprise to uncover universal laws governing human activity, positing instead that in the realm of human phenomena variation and difference were inescapable and might in fact be preferable to uniformity (Berlin 2000, p. 16). They further eschewed controlled experimentation, quantification, and measurement, which they argued held little potential to advance the ultimate goal of not simply understanding but ameliorating the human condition (see Cole 1996). This notion of a non-reductive, holistic human science continues to be influential in the social sciences and humanities, and for Vygotsky and Luria it aligned well with their dialectical approach to understanding the human mind through change.

## Dialectics of Development

As explained earlier, the move to render psychology scientific led to suggestions that the field be divided into two realms, one that relied on heuristics and the other that employed natural science methods. Vygotsky and his colleagues strongly resisted such proposals, working instead to create a theoretically unified psychology, and their initial step was to overcome a number of dualisms inherited from Cartesian logic that they believed impeded progress in the field (Bakhurst, 2007). While these are crucial to fully understanding Vygotsky's research program, they are not central to our concern with DA (although interested readers should refer to van der Veer & Valsiner, 1991, and Lantolf & Thorne, 2006). I will however briefly discuss one problem Vygotsky faced that is relevant to DA and that is illustrative of his theoretical design: the human-world relation.

Valsiner and van der Veer (2000, p. 6) explain that psychology has long struggled to conceptualize the relation of human beings to the world and to resolve the problem that the mind, a non-physical representational structure, could understand the physical world. Ilyenkov (1977) goes further, suggesting that the human-world, or mind-body, dualism is the fundamental problem of philosophy and its resolution is central to understanding human existence. He points out that by positing that mind and body are not merely different entities or substances but are in fact opposite, Descartes and his followers created a dualism for which they had no resolution, save invoking supernatural authority: mind and body were both creations of God, and so it is through divine will that they are interrelated (Ilyenkov 1977, p. 26). In contrast, a dialectic mode of thought, according to Ilyenkov, offers an approach to the problem that does not require divine intervention. It was the Dutch philosopher Baruch Spinoza who first suggested that there was no solution to Descartes' dualism simply because the problem had been incorrectly formulated. In Spinoza's view, there was no need to relate mind to body because they were not two separate objects in the first place but one:

There are not two different and originally, contrary objects of investigation – body and thought – but only *one single object*, which is the *thinking body of living, real man* (or other analogous being, if such exists anywhere in the Universe), only considered from two different and even opposing aspects or points of view. Living, real thinking man, the sole thinking body with which we are acquainted, does not consist of two Cartesian halves – 'thought lacking a body' and a 'body lacking thought'... and one cannot in the end model a real thinking man from two equally fallacious abstractions (Ilyenkov 1977, p. 31, italics in original).

Importantly, the mind-body relation is only an insoluble problem if conceptualized in a dualistic rather than dialectical manner. Bakhurst (1991) argues that the crucial difference between these views is that while the former is above all interested in how two things, *a* and *b*, are different from one another, the latter posits a *unity of opposites* whereby the differences between *a* and *b* constitute the strength of their interrelationships and enable them, together, to form a unified whole (p. 139). It is this integrated whole, this synthesis of *a* with *b*, that is the object of interest in dialectics (see also Acton 1967; Novack 1971). Bakhurst (1991) offers a compelling illustration of this perspective by likening a dialectic-



cal unity to the ability to listen to music rather than simply hearing individual notes. Only by appreciating the piece of music as a whole can one understand why specific notes are played in prescribed ways and why pauses are necessary in particular places (pp. 172-173).

Returning to the mind-body problem, Vygotsky (1987) broke with psychologists who posited that mental functions are purely biological in nature (i.e., genetically endowed), but he also took issue with those who maintained that mental functions are derived from the environment and consequently preserve a physical essence (Valsiner & van der Veer 2000, pp. 6-7). Proceeding from a dialectical perspective, Vygotsky argued a nature-nurture way of framing the debate, insisting that neither biology nor the environment can fully explain the human mind but that these must be taken together as a dialectical unity (Vygotsky, 1987). The brain, a biological organ, is a necessary precondition to the development of mental processes but the processes themselves arise as individuals use culturally created artifacts and engage in activities with others (Valsiner & van der Veer 2000, p. 364). Physical and symbolic artifacts function as mediating tools, enabling humans to gain control over their mind, as for example when mnemonic devices regulate simple memory, knowledge of things in the world shapes perception or attracts and maintains attention, and verbal reasoning helps individuals to think through problems (Kozulin 1998).

Vygotsky's dialectical response to the mind-body problem is expressed in the axiom that cognitive functions actually appear twice – first on the *interpsychological plane* of joint activity among individuals acting with cultural artifacts and later on the *intrapsychological plane* as new functions with which individuals regulate their own behavior (Vygotsky 1978, p. 57). In Vygotskian theory, this process of ingrowing, or internalizing cultural forms of mediation, is development (Lantolf 2003). His celebrated concept of the Zone of Proximal Development (ZPD) is at once an expression of the interrelationship of mediation, internalization, and development as well as a practical framework for pedagogical interactions. The ZPD brings to light abilities that have not yet fully developed, functions that are still in the process of being internalized. As such, the ZPD offers a target, albeit a moving one, for teaching: instruction that is beyond an individual's current level of development and that is oriented to emerging abilities may maximally impact and guide development. At the same time, the ZPD is revealed not by observation of solo performance but through collaboration between teachers and learners, in which various forms of mediation are offered by the teacher/mediator and responded to (accepted/rejected/modified) by the learner. A practical consequence, then, of Vygotsky's dialectical approach to human development is that it integrates teaching and assessment in a single activity in which mediation is used to uncover a learners' ZPD while this interaction in itself moves the ZPD forward.

DA practitioners have devised a number of approaches to co-constructing ZPDs with individuals, and differences among these approaches are in part attributable to the diversity of contexts, problems, and populations to which DA principles have been applied. For example, much of the early DA work was conducted in the areas of intelligence testing and special education, which concerned the diagnosis of learning problems (Burdock 1968; Feuerstein, Rand & Hoffman 1979). Given these assessment contexts, DA was often limited to the clinical work of school psychologists and remained divorced from everyday teaching and learning. Haywood and Lidz's (2007) recent overview of DA research suggests that this trend is continuing but that there are also important exceptions, most notably the work associated with Reuven Feuerstein's International Center for the

Enhancement of Learning Potential (see Feuerstein, Falik, Rand & Feuerstein 2003). In the L2 domain, the growing DA research literature has remained squarely focused on the classroom as a way of bringing teaching and assessment together to support learner development (e.g., Ableeva 2008; Poehner 2008; Summers in preparation). The next section presents some brief examples of DA with L2 learners before returning to the matter of validity.

### Dynamic Assessment and L2 Learners

Poehner (2007, 2008) designed a DA program for undergraduate university learners of L2 French. Learners were at an advanced level (fourth-year) of language study and were enrolled in a course focused on oral communication. At the outset of the course, learners met individually with a mediator and were asked to orally compose narratives in the L2. Learners watched brief video clips from the film *Nine Months*, a lighthearted comedy from the Nineteen-nineties featuring well-known actors Hugh Grant, Julianne Moore, and Robin Williams. The clips were chosen because they offered rich material for student narratives: humorous situations, unexpected events, character motivations and relationships, and sequences of related occurrences.

During the initial diagnostic, learners were asked to narrate two short clips from the film. The first they narrated on their own, but for the second they were offered dialogic support from the mediator. Individuals' performance on the two tasks formed the basis of an ongoing enrichment program aimed at remediating problems identified during the diagnosis by continuing to collaboratively engage in storytelling, always prompted by clips from movies. In keeping with DA's dialectical approach to teaching and assessment, all mediator-learner interactions attempted to co-create a ZPD regardless of whether it was the initial or the final meeting. That is, unlike a more traditional formative assessment cycle in which an assessment is followed by additional instruction that in turn leads to another assessment, all sessions involved simultaneous teaching and assessment as the mediator interpreted learner moves, offered support, and cooperated with learners to help them stretch beyond their current functioning. Indeed, during later sessions the narration tasks were deliberately rendered more challenging by employing increasingly complex prompts (including powerful scenes from dramatic films as well as literary passages) in order to ascertain the degree to which learners could extend their emerging abilities beyond the immediate context, a feature of DA known as *transcendence* (see Poehner 2007).

The examples that follow concern two learners, Nancy (N) and Elaine (E). N experienced difficulty selecting object pronouns and correctly placing them in constructions while E struggled to employ present perfective and imperfective aspect to situate events in her narrative. Both issues typically pose challenges for English-speaking learners of French. In the case of object pronouns, this is because French and English follow different syntactic patterns. With regard to verbal aspect, this is marked differently in the two languages, with French employing the *passé composé* (present perfective) or the *imparfait* (present imperfective) to either emphasize the completed or ongoing aspect of actions, events, and states of being. In narratives, the imperfective is generally used to express background information while perfective aspect foregrounds information that moves the events of the narrative forward in time. Both interactions below are taken from the learners' initial DA sessions during which the mediator (M) attempts to work within their ZPD. These

excerpts of DA cooperative dialoguing illustrate unique features of the approach, including insights into the dynamics of learner development typically not afforded by NDA.

#### A Moving Target

Although the ZPD offers a particularly powerful way of approaching learner development, it is clearly not a quick or easy set of techniques that will lead to promised results. In fact, the DA work reported by Feuerstein and his colleagues offers intensive cognitive intervention plans for individuals with learning disabilities that in extreme cases extend over months or even years.<sup>6</sup> Nonetheless, Vygotzky's own research provided evidence of development that can occur in a short period time, even during a single interaction. As Wertsch (1985, p. 55) explains, one of Vygotzky's important criticisms of experimental research in psychology was that it frequently discarded the most interesting data, which occurred during the time in which participants came to understand the conditions, aims, and demands of the experiment. Vygotzky (1978, p. 68) charged that even these immediate responses represent a developmental process, which Wertsch (1985, p. 55) refers to as the microgenetic domain of development.

In (1) below, N is narrating a scene from *Nine Months* in which two characters, Rebecca and Samuel, are arguing and she attempts to relate that Rebecca thought Samuel was accusing her of something.

- (1)
1. N: ...et à ce moment Rebecca a pensé que Sam a accusé elle\* a accusé elle\* de ne pas (...)  
...and at that moment Rebecca thought that Sam accused her\* accused her\* of not (...)
  2. M: a accusé elle\*  
accused her\*?
  3. N: (...) elle a lui accusé\*? I guess that would be wrong a accusé elle\* can you say that?  
(...) accused her\*?
  4. M: well usually you would use the elle as an object right
  5. N: right
  6. M: so like he accused [her
  7. N: her
  8. M: so you'd have to insert it like a direct object
  9. N: so a elle accusé\* a elle accusé\*?  
accused her\* accused her\*?
  10. M: well remember the objects usually go before the verb

<sup>6</sup> Here the interested reader may refer to the Instrumental Enrichment program designed by Reuven Feuerstein (Feuerstein, Rand, Hoffman & Miller 1980).

11. N: but if it's in past tense is it after the avoir and then before the past participle
12. M: actually they go before the before the avoir
13. N: but I'm saying it's Sam accused her
14. M: right
15. N: so how would I...
16. M: it's just the order of the words would be different in French so instead of Sam accused her it would be Sam her accused
17. N: Sam her accused so uh Sam elle a accusée\*?
18. M: except the elle would be switched to I'a accusée
19. N: I'accusé\*! Uh I'l'a accusée  
accused her
20. M: I'a accusée
21. N: (laughs) okay so Sam I'a accusée (...) okay so uh Rebecca a pensée...  
Sam accused her (...) okay so uh Rebecca thought...

Note in the above exchange that N appears even at the outset to be aware that the construction *a accusé elle* is problematic. Her response to M's repetition of her utterance is not to confirm it but rather to reformulate it as *elle a lui accusé*. This, too, however is incorrect as N herself points out. She returns in line 3 to her original construction but poses it as a question to M, seeking his approval. Even at this point in the interaction, where M has done very little, a good deal has been learned about N's control over French object pronouns and their placement in perfective constructions. To be sure, she does not have adequate control to produce the needed structure on her own, a fact that would have been noted in a non-dynamic procedure. However, dialoguing with the learner reveals that she knows she has made an error and, more specifically, that the error concerns both the selection and placement of the pronoun (she switches to *lui* and proposes it in relation to the participle *accusé*), something that would not emerge in traditional assessment.

Even greater insights into N's abilities are gained as the interaction continues. M's clue in line 8 that the pronoun functions as a direct object leads her to move *elle* from a final position to a place between the auxiliary and the past participle. In line 11, N contests M's suggestions, arguing that her choice reflects the appropriate pattern of clitic placement in French, and it is only after further discussion that she realizes that the French present perfective is composed of two verbal elements and that clitics must precede them both (in other verbal constructions, French does not use an auxiliary and so the pronoun would simply precede the conjugated form of *accuser*). M's support becomes increasingly explicit, and ultimately N does not produce the construction herself but only repeats it and inserts it into her narrative once M has provided it.

From this, one might conclude that producing the task is beyond N's present capabilities. Simply repeating the utterance is not, in itself, evidence that the learner is

developing the ability to control this feature of the language. However, it is equally evident from the interaction that N has some knowledge of clitic placement in French and, further, that she is aware that her constructions were not correct. Nancy's ability to employ French clitics correctly in spoken discourse is clearly not fully developed but it is forming. In fact, her repetition of the appropriate construction in line 21 may constitute a form of imitation, which Vygotzky (1987) viewed as an important indicator of the ZPD. For Vygotzky, imitation is far more than mere parroting or mimicry as it suggests an understanding of the meaning and intentions associated with particular behaviors. For instance, a child's imitation of an adult performing a simple task such as opening a can of tomatoes requires that the child understand that something is in the can that must be removed, that a tool such as a can opener is required to accomplish the task, and that this tool must be applied in a specified manner in order to be effective (see Tomasello 2003). Only at a certain stage in the child's development will s/he attend to all these dimensions of the adult's behavior and imitate them. From this, Vygotzky concluded that individuals are only able to imitate those functions that lay within their ZPD (Vygotzky 1987, p. 209). In N's case, her performance later during this same interaction offers evidence that her repetition of the correct construction had an imitative quality and that control over French clitics was within her ZPD.

In (2) below, N is explaining that two other characters in the film, Sean and Christine, have ended their relationship. She initially tries to express this idea by relating that they are no longer together, but M prompts her in line 2 to frame the event as it was portrayed in the video clip: Sean is troubled because Christine left him. This requires use of an object pronoun in a perfective construction just as before:

- (2)
1. N: ...so Sean a dit que il que Christine et il ne sont pas [together]  
Sean said that he that Christine and he are not [together]
  2. M: one thing you could say is that she left him
  3. N: she left him ah (...) I'm trying to think of how you would say she left
  4. him uh (...) elle est partie de sa vie (laughs)  
she departed his life
  5. M: (?)
  6. N: I don't know how you say she left him
  7. M: you can use the verb quitter  
to leave
  8. N: oh quitter! so elle l'a elle l'a quitté?  
oh to leave! so she him she left him?
  9. M: voilà

N responds to M's prompt that she emphasize Christine's act of leaving with the somewhat

melodramatic *elle est partie de sa vie* in line 4. Her laughter suggests that she realizes how unusual this sounds, and in line 6 she turns to M for help. Following the same approach of beginning with implicit mediation and becoming more explicit as needed, M offers the lexical item *quitter* (to leave), and this proves to be the only support N requires. In line 8 she appears to recognize this infinitive and then produces a present perfective construction with the correct object pronoun appropriately placed: *elle l'a quitté*. In effect, it was lexical knowledge rather than clitic placement that proved problematic for N at time (2).

N's contributions to the performance during (2) indicate that she had begun to internalize the support offered to her earlier and that she was becoming capable of employing it in a more autonomous manner. This is not to say that the observed changes may be solely attributed to the mediator's earlier intervention, but that the initial dialoguing targeted an ability that was forming (i.e., it was within N's ZPD) and so was able to impact its development. Even in (2) N still was not yet able to perform completely independently, and a non-dynamic procedure may have simply noted that in neither instance was she able to 'get it right.' However, such a diagnosis would obscure important differences in her contributions to the activity and her true level of development – including her emergent capabilities – would be missed. The next examples we will consider concern the potential for dialoguing to bring to light underlying problems, thereby expanding the evidential basis for interpreting learner abilities.

#### Beyond Observable Behavior

While E narrated these same video clips, M noted a different phenomenon: an almost total reliance on the present perfective. E recounted her narrative as a series of events, and the lack of imperfect aspect meant that none of the events were moved to the foreground or background vis-à-vis the others but that the retelling followed a list-like structure. Indeed, only the verb *être* (to be) appeared in the imperfect, and this verb is traditionally presented in grammar books as a primary vehicle for background descriptions (e.g., we were in Paris, they were tired). Moreover, verbs were used in the perfective even in instances that clearly required the imperfective, as in "*quand Christine était avec lui elle a voulu\* avoir des enfants* (when Christine was with him she wanted [perfective] to have children)." In this particular case, when M questioned her choice of verbal aspect E simply replied, "*je ne peux pas expliquer c'est la façon dans laquelle je parle* (I can't explain that's the way I speak)." At one point in their interaction M offered the learner the opportunity to renarrate a portion of the video clip and once again prompted her to reflect on her use of aspect:

- (3)
1. M: go ahead you can go back through it now real quick in French just the
  2. part where you were setting it up?
  3. E: C'étaient dans la voiture rouge et ils ont ils ont conduit. Rebecca a  
It was in the red car and they they drove
  4. je pense que j'ai dit elle a compté dans la calendrier

*I think that I said she counted on the calendar*

5. M: uh huh elle a compté et ils ont conduit so passé composé and then how she counted and they drove
6. come passé composé there cause [you used it again there
7. E: I don't know it just it just is that a good explanation? Because it just
8. sometimes that's how you say it?

E's response in lines 7-8 suggests either a reluctance or inability to verbalize the reasons behind her approach to framing these events in the narrative. Through further discussion, M is able to ascertain that it is not that E is simply unaccustomed to thinking in depth about how perfective and imperfective aspect may be used to describe the past, but rather that her understanding of this feature of the language is not well developed:

9. M: ...and if you were to go back and do it now or to write it as you said?
10. E: I would probably use the imperfect
11. M: oh instead of passé composé?
12. E: (...) yeah. If I was writing it I might have just picked one of the two
13. and then stuck with it for the whole thing.
14. M: one of the two? Like either imperfect or passé composé?
15. E: yeah
16. M: and stuck with it for everything?
17. E: yeah for the most part of it.

Despite her earlier statement that the perfective had a sort of intuitive appeal for this part of the narrative, she remarks in line 10 that if she were to redo the task she would opt for the imperfective. This reversal could indicate E's awareness on some level that either aspect may be used to talk about the past, but her subsequent remarks make little sense. Outside of the world of creative or experimental literary endeavors, it is difficult to imagine why (and indeed, how) one would compose a narrative employing only one aspect or the other. Furthermore, this begs the question as to how one decides which aspect to use in the narrative. M inquires about this and E explains as follows:

18. E: Passé composé being the action it happened once either it happened
19. once or it happened completely and it's over a habitual action where it
20. keeps on going or it's still going uh it's still going on. I don't know if that's

right or not...

E undoubtedly has some knowledge of the passé composé – imparfait distinction, and appears to be attempting to follow certain normative rules that she likely encountered through formal study of the language (e.g., rules connecting imperfective aspect to habitual actions). She only partially recalls these rules, and it is clear that she does not fully understand them. For example, all the rules are linked to the perfective, and her statement about actions that are “still going on” is somewhat ambiguous (is it ongoing at the time of narration, as in “I have been a student for three years,” with the implication that I am still a student?). In addition, it is not evident given her performance whether she is in fact basing her decisions on an application of these rules or if she simply invokes them at this point in order to offer some explanation to M. At any rate, E's interaction with M allows for a very different picture of her abilities to emerge than the impression one might get from simply observing her performance. In particular, her control over verbal aspect is not yet adequate to enable her to appropriately discuss the past in French.

### Discussion

The preceding examples should give a sense of the form L2 DA interactions might take, and more important, should make clear that establishing the validity of such interactions is qualitatively different from other forms of assessment. Validation here is not a matter of argumentation to support generalizations about individuals' underlying abilities because DA is concerned with joint (rather than solo) activity intended to move beyond the level of independent performance. The quality of mediator-learner dialoguing is constantly in flux as problems arise and support is negotiated, and so rather than controlling all variables to obtain an accurate measure of where abilities lie at a given point, DA seeks to create conditions for development to take place. The central argument to establish validity is not whether the assessment produced appropriate measures of the abilities in question but to what degree it supported learner development. Of course, considering ‘what worked’ is fundamental to classroom practice, but a teaching-assessment/talesic offers the possibility for more focused reflection, in which development remains in the foreground at all times.

Having said this, some ideas from psychometric test validation may be encompassed by this reflection on development, but they will take on new meanings. Messick's unified model of validity is an excellent point of departure for such re-envisioning as it provides anchoring points to which teachers might connect their reflections on classroom interactions and their impact on learner development. Recall that in Figure 1, Messick proposed constructs, evidence, interpretations, and consequences as the foundations of his validity framework. In light of the preceding discussion of L2 DA examples, we can consider how Messick's pillars might be reframed from a development-oriented perspective.

Beginning with the matter of constructs, or abilities to be assessed, psychometric testing emphasizes appropriate modeling and elimination of irrelevant variance. From a development-oriented perspective, the focus is on changes in learner abilities. This may come as learners at one point in time are capable of performing independently what previously had only been possible with support from someone else. However, learner develop-

ment may also manifest itself in changes in the kinds of mediation they require or in how they respond to mediation. For example, Nancy required extensive mediation in order to produce the construction *Sam l'a accusé*. Her difficulty in selecting an appropriate object pronoun and inserting it in a perfective construction prompted the mediator to ultimately reveal the solution. Later, when expressing the idea that Christine left Sean, the learner was confronted with a similar problem but in this case required only lexical support (the verb *quitter*). In neither case was Nancy able to produce the necessary constructions entirely on her own, but there was a clear and important shift in the quality of support required. This does not mean that Nancy had complete control over clitics but, as already explained, it does demonstrate the potential impact of cooperative dialoguing on learner development. Rather than treating this as a source of variance that invalidates our interpretation of Nancy's abilities, this kind of change in performance is the desired outcome because it suggests, as Vygotsky would say, that the learner's abilities are "ripening".

Regarding the evidential basis for interpretations, mediator-learner interactions enable us to look beyond whether learners are able to carry out tasks successfully on their own. In the examples with Nancy, these interactions pointed to changes in the mediation she needed; in Elaine's interactions with the mediator, the interactions helped to externalize cognitive processes, rendering them visible for examination and intervention. It was only through discussion of Elaine's choices that the source of the problem became clear: she did not have an adequate understanding of verbal aspect to appropriately use the *passé composé* and imparfait to frame events in her narratives. This has important implications for diagnosis (and remediation) needed to promote development, as it is simply not the case that Elaine made a conscious choice to construct her narrative as a series of events or even that she made occasional mistakes but otherwise had good control over the present perfective. Elaine's verbalizations of her understanding of aspect were crucial to the diagnosis, and as with the type of mediation Nancy required, it is difficult to imagine that this evidence of abilities could have been so readily, if ever, obtained through other assessments.

Interpreting evidence of learner abilities is rendered more complex by virtue of the fact that the evidence itself is dynamic rather than stable. That is, a teaching-assessment dialectic requires that in addition to interpretations of assessment outcomes after the activity is completed, there is the perhaps more important level of moment-by-moment interpretation *during* the interaction as the mediator continually works to calibrate and negotiate support with the learner. This emergent interpretation is vital to the teaching-assessment dialectic because appropriate mediation depends upon understanding the type of assistance the learner needs at a given moment – mediation that is overly implicit will not be helpful to the learner and mediation that is overly explicit (e.g., simply providing the answer at the first sign of difficulty) undermines the diagnostic potential of the interaction because one does not know the type of support the learner really needs and where s/he is developmentally. In Example (1), the mediator did not immediately offer Nancy the needed construction nor did he even explicitly alert her to the fact that she had committed an error. Rather, he began by simply repeating her utterance and then confirming that she intended to use *elle* (she) as an indirect object. This exchange provided Nancy with the opportunity to correct her own mistake (which she was unable to do) and it also made apparent to both mediator and learner the problem to be addressed. The mediator's subsequent moves, which included metalin-

guistic clues, translation of the construction into English, and ultimately providing the structure itself, all took account of the learner's responsiveness while encouraging her to take on as much responsibility for the performance as she was able. In other words, the dialectic demands that at every step interpretation is integrated with intervention.

This way of approaching teaching and assessment has similar implications for thinking about assessment consequences. As explained, assessment consequences – for individuals, institutions, and society – have emerged as a topic of great concern as research has pointed increasingly towards the political agendas behind many assessment initiatives and the adverse effects of mandated assessments on educational systems. In DA, however, the integration of classroom teaching and assessment has the intentional (and positive) consequence of leading the development of all learners. Indeed, the principal question for classroom practitioners must be to what extent their teaching-assessment activities advance learner development. This should not be confused with so-called positive washback, or high systemic validity, in which a test is introduced with the intention of improving learning by focusing teachers and students on meeting specified curricular goals and standards. Even if one were able to engineer a test for positive washback (an assumption for which, to the best of my knowledge, no convincing evidence has yet been presented), the educational enterprise would remain an activity oriented towards test preparation with the test controlling the kinds of knowledge and understanding that are valued. The development-orientation proposed here is not immediately concerned with obtaining correct answers or improving scores. Internalizing mediational means may lead to improved performance in the traditional assessment sense but, more importantly, it enables individuals to engage in communicative activity in more agentive ways and this goes well beyond mastery of test items.

## Conclusion

This paper began from the premise that classroom-based assessment often carries high stakes for teachers and learners but that the important matter of validating these assessments has not been fully addressed in the literature. Previous discussions of validity in classroom assessment raised the question of whether psychometric approaches to test validity are an appropriate framework for the classroom or if other ways of thinking about validity are needed. I have argued in favor of the view that psychometric testing and classroom-based formative assessment are fundamentally distinct activities with differing goals and methods. The psychometric paradigm is modeled after experimental research in the natural sciences, and this way of approaching assessment imposes limitations and restrictions that need not be imported into the classroom, where the purpose is not to measure abilities but to support teaching and learning. I have proposed that the work of Vygotsky and his colleagues offers a principled approach to integrating teaching and assessment as well as a theoretical framework for reflections on the validity of this activity. The relevance of Vygotsky's dynamic model of assessment to the classroom is that assessment is conceived of as a way of helping individuals move beyond their current capabilities. From this perspective, understanding learners' current level of independent performance is of interest only as a point of departure for helping the learner construct a future. In this way, classroom assessment, conceived as intervention, is not about classifying or cat-

egorizing individuals but may be understood as a pedagogical form of activism designed not simply to improve the odds of success for some learners to change the odds for all.

## References

- Ableeva, R. (2008). The effects of dynamic assessment on L2 listening comprehension, in J. P. Lantolf and M. E. Poehner (eds.), *Sociocultural theory and the teaching of second languages*. London: Equinox.
- Acton, H. B. (1967) Dialectic materialism. In P. Edwards (ed.) *The Encyclopedia of Philosophy* Vol. 2 389-397. New York: Macmillan and the Free Press.
- Alderson, J.R., & D. Wall. (1993). Does washback exist? *Applied Linguistics* 14: 115-29.
- Anastasi, A. & S. Urbina. (1996). *Psychological Testing. Seventh Edition*. Prentice Hall: Upper Saddle River, NJ.
- Bachman, L. F. (2000). Modern language testing at the turn of the century: assuring that what we count counts. *Language Testing* 17: 1-42.
- Bakhurst, D. (2007) Vygotsky's demons. In H. Daniels, M. Cole & J. V. Wertsch (eds.) *The Cambridge companion to Vygotsky*. Cambridge: Cambridge University Press, pp. 50-76.
- Berlin, I. (2000). *Three critics of the Enlightenment. Vico, Hamann, Herder*. In H. Hardy (Ed.). Pimlico: London.
- Block, D. (1996). Not so fast: Some thoughts on theory culling, relativism, accepted findings and the heart and soul of SLA. *Applied Linguistics* 17: 63-83.
- Budoff, M. (1968). Learning potential as a supplementary testing procedure. In *Learning Disorders. Vol. 3*. J. Hellmuth (Ed.). Seattle, WA: Special Child.
- Cheng, L. (2005). *Changing Language Teaching through Language Testing: A Washback Study*. Cambridge: Cambridge University Press.
- Cheng, L., Y. Watanabe, and A. Curtis (Eds.). (2004). *Washback in Language Testing: Research Contexts and Methods*. Mahwah, NJ: Lawrence Erlbaum.
- Cole, M. (1996). *Cultural Psychology. A Once and Future Discipline*. New York: Bradford Books.
- Cronbach, L. J. (1990). *Essentials of Psychological Testing. 5th Edition*. New York: Harper & Row Publishers.
- Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin* 52: 281-302.
- Daniels, H. (1993). (ed.). *Charting the agenda: Educational activity after Vygotsky*. London: Routledge.
- Danziger, K. (1990). *Constructing the subject: Historical origins of psychological research*. Cambridge: Cambridge University Press.
- Danziger, K. (1997). *Naming the Mind: How psychology found its language*. Thousand Oaks, CA: Sage.
- Davies, A. (2003). Three heresies of language testing research. *Language Testing* 20: 355-368.
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.
- Feuerstein, R., L. Falik, Y. Rand, & R.S. Feuerstein (2003). *Dynamic Assessment of Cognitive Modifiability*. Jerusalem: ICELP Press.
- Feuerstein, R. Y. Rand, M. B. Hoffman, and R. Miller. (1980). *Instrumental Enrichment*. Baltimore, MD: University Park Press.

- 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.
- Gadamer, H. G. (1975). *Truth and method*. New York: Seabury.
- Gipps, C. V. (1994). *Beyond testing: Towards a theory of educational assessment*. London: Falmer Press.
- Good, S. R. (2000). A. R. Luria: A humanistic legacy. *Journal of Humanistic Psychology* 40: 17-32.
- Gould, S.J. (1996). *The Mismeasure of Man. Revised and Expanded*. New York: W.W. Norton & Co.
- Gregg, K. (1993). Taking explanation seriously; or, let a couple of flowers bloom. *Applied Linguistics* 14: 276-294.
- Hanson, F.A. (1993). *Testing testing: Social consequences of the examined life*. Berkeley: University of California Press.
- Haywood, H. C. & C. S. Lidz. (2007). *Dynamic assessment in practice. Clinical and educational applications*. Cambridge: Cambridge University Press.
- Ilyenkov, E. V. (1977). *Dialectical logic*. Moscow: Progress Publishers.
- Johnson, K.E., S. Rehr-Jordan, & M.E. Poehner. (2005). The TOEFL Trump Card: An Investigation of Test Impact in an ESL Classroom. *Critical Inquiry in Language Studies* 2 (2): 71-94.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin* 12: 527-535.
- Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement* 38: 319-342.
- Karpov, Y.V. (2005). *The Neo-Vygotskian approach to child development*. Cambridge: Cambridge University Press.
- Kozulin, A. (1998). *Psychological tools: A sociocultural approach to education*. Cambridge, MA: Harvard University Press.
- Kozulin, A., B. Gindis, V. S. Ageyev & S. M. Miller. (2003). (eds.) *Vygotsky's educational theory in cultural context*. Cambridge: Cambridge University Press.
- Lantolf, J. P. (1996). Second language theory building: Letting all the flowers bloom! *Language Learning* 46: 713-749.
- Lantolf, J. P. (2000). Rethinking the relevance of experimental research for second language classroom. In S. Salaberrí (ed.). *Linguística aplicada a la enseñanza de lenguas extranjeras*. Almería: University of Almería Press.
- Lantolf, J. P. (2003). Intrapersonal communication and internalization in the second language classroom. In A. Kozulin, B. Gindis, V. S. Ageyev & S. M. Miller (eds.). *Vygotsky's educational theory in cultural context*. Cambridge: Cambridge University Press.
- Lantolf, J.P. & M.E. Poehner. (2004). Dynamic Assessment: Bringing the Past into the Future. *Journal of Applied Linguistics* 1: 49-74.
- Lantolf, J.P. & M.E. Poehner. (2007). Dynamic Assessment. In *Encyclopedia of Language and Education Volume 7. Language Testing and Assessment* (E. Shohamy, Ed., N. Hornberger, Gen. Ed.). Springer Publishing.
- Lantolf, J. P. & M. E. Poehner (2008). *Sociocultural theory and the teaching of second languages*. London: Equinox.
- Lantolf, J.P. & S.L. Thorne. (2006). *The Sociogenesis of Second Language Development*. Oxford: Oxford University Press.
- Lave, J. & E. Wenger. (1991). *Situated learning. Legitimate peripheral participation*. Cambridge: Cambridge University Press.
- Leung, C. (2007). Dynamic assessment: Assessment for and as teaching? *Language Assessment Quarterly: An International Journal* 4: 257-278.
- Long, M. H. (1993). Assessment strategies for SLA theories. *Applied Linguistics* 14: 225-249.
- Luria, A. R. (1961). Study of the abnormal child. *American Journal of Orthopsychiatry*. A *Journal of Human Behavior* 31: 1-16.
- Luria, A.R. (1979). *The Making of Mind: A Personal Account of Soviet Psychology*. Cambridge, MA: Harvard University Press.
- Lynch, B. K. (2001). Rethinking assessment from a critical perspective. *Language Testing* 18: 351-372.
- McNamara, T. F. (2001). Language assessment as social practice: challenges for research. *Language Testing* 18: 333-349.
- McNamara, T. (2006). Validity in language testing: The challenge of Sam Messick's legacy. *Language Assessment Quarterly* 3: 31-51.
- McNamara, T. & C. Roever. (2006). *Language testing. The social dimension*. Hoboken, NJ: Wiley-Blackwell.
- 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.A. (1989). Validity. In *Educational Measurement Third Edition*. R.L. Linn (Ed.). New York: American Council on Education.
- Messick, S. (1995). Validity of psychological assessment. Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist* 50, (9): 741-749.
- Mislevy, R. J. (1996). Test theory reconceived. *Journal of Educational Measurement* 33: 379-416.
- Moss, P.A. (2003). Reconceptualizing validity for classroom assessment. *Educational Measurement: Issues and Practice* 22, (4): 13-25.
- Novack, G. (1971). *An introduction to the logic of Marxism*. New York: Pathfinder Press.
- Poehner, M.E. (2007). Beyond the Test: L2 Dynamic Assessment and the Transcendence of Mediated Learning. *The Modern Language Journal* 91: 323-340.
- Poehner, M.E. (2008). *Dynamic Assessment: A Vygotskian Approach to Understanding and Promoting Second Language Development*. Berlin: Springer Publishing.
- Polkinghorne, D. (1983). *Methodology for the human sciences. Systems of inquiry*. Albany, NY: State University of New York Press.
- Rampton, B. (1995). Politics and change in research in applied linguistics. *Applied Linguistics* 16: 233-256.
- Ratner, C. (2002). *Cultural Psychology. Theory and Methods*. New York: Plenum Press.
- Ratner, C. (2006). *Cultural Psychology. A Perspective on Psychological Functioning and Social Reform*. Mahwah, NJ: Lawrence Erlbaum Associates.



- Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing*, 18(4), 429-462.
- Rea-Dickins, P., & Gardner, S. (2000). Snarles and silver bullets: Disentangling the construct of formative assessment. *Language Testing*, 17(2), 217-244.
- Sacks, P. (1999). *Standardized Minds. The High Price of America's Testing Culture and What We Can Do To Change It*. Cambridge, MA: Perseus Publishing.
- Salomon, G. (ed.) (1996). *Distributed cognitions: Psychological and educational considerations*. Cambridge: Cambridge University Press.
- Shohamy, E. (2001). *The Power of Tests: A Critical Perspective on the Uses of Language Tests*. New York: Longman.
- Sternberg, R.J. and E.L. Grigorenko. (2002). *Dynamic Testing. The Nature and Measurement of Learning Potential*. Cambridge: Cambridge University Press.
- Sullivan, E. V. (1984). *A critical psychology. Interpretation of the personal world*. New York: Plenum.
- Summers, R. (in preparation). *Computer Mediated Dynamic Assessment: Toward an Online Model of Dialogic Engagement*. Unpublished doctoral dissertation, University of South Florida.
- Teasdale, A., & Leung, C. (2000). Teacher assessment and psychometric theory: A case of paradigm crossing? *Language Testing*, 17(2), 163-184.
- Tomasello, M. (2003). *Constructing a language. A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Torrance, H. (1995). Teacher involvement in new approaches to assessment. In H. Torrance (ed.), *Evaluating authentic assessment*. Buckingham: Open University Press, pp. 44-56.
- Torrance, H. & J. Pryor. (1998). *Investigating formative assessment: teaching, learning and assessment in the classroom*. Buckingham: Open University Press.
- Valsiner, J. & R. van der Veer. (2000). *The social mind. Construction of the idea*. Cambridge: Cambridge University Press.
- van der Veer, R. & J. Valsiner. (1991). *Understanding Vygotsky*. Oxford: Blackwell.
- van Lier, L. (2004). *The Ecology and Semiotics of Language Learning. A Sociocultural Perspective*. Boston: Kluwer Academic.
- Vygotsky, L.S. (1978). *Mind in Society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Vygotsky, L. S. (1987). Thinking and speech. In R. W. Rieber & A. S. Carton (eds.), *The collected works of L. S. Vygotsky. Vol. 1. Problems of general psychology* (pp. 39-285). New York: Plenum.
- Vygotsky, L. S. (1997). The Historical Meaning of the Crisis in Psychology: A Methodological Investigation. In *The Collected Works of L. S. Vygotsky. Vol. 3. Problems of the Theory and History of Psychology* (pp. 233-344). R. W. Rieber & J. Wollcock (Eds.). New York: Plenum.
- Vygotsky, L. S. (1998). The Problem of Age. In *The Collected Works of L. S. Vygotsky. Vol. 5. Child Psychology*. R. W. Rieber (Ed.). New York: Plenum.
- Wenger, E. (1998). *Communities of practice. Learning, meaning, and identity*. Cambridge: Cambridge University Press.
- Wertsch, J. V. (1985). *Vygotsky and the social formation of mind*. Cambridge, MA: Harvard University Press.
- Wells, G. & G. Claxton. (2002). *Learning for life in the 21<sup>st</sup> Century: Sociocultural perspectives on the future of education*. Oxford: Blackwell Publishing.
- Yngve, V. (1996). *From grammar to science: New foundations for general linguistics*. Amsterdam: John Benjamins.