

Myth Busting and Metric Making: Refashioning the Discourse about Development *Excursus for Integral Leadership Review*

Zachary Stein¹
Developmental Testing Service, LCC
Harvard University Graduate School of Education

Executive summary: This paper is about the use we make of developmental metrics and models in the Integral Community. Motivated by broad concerns about growing markets for psychological technologies I propose the need for discourse about quality control parameters, focusing on the construction and deployment of developmental assessment systems. When it comes to developmental metrics and models **I suggest that two myths loom large and that they need to be busted** if we want to move forward responsibly. The big take home here is that **the most popular approaches are not the most preferable.**

Discussing *the myth of the given* raises concerns about our *methods* and how we determine the relative validity of our various metrics and models. I suggest that we are systematically misunderstanding the nature of developmental approaches because we are not focusing on how we build usable knowledge about human development. The point: **we need to be more concerned about the claims we make, and start putting our methods where our mouths are.** I ground this discussion by looking at an approach that has jettisoned this myth because it is systemically monitoring the validity of the methods in use and publishing the results of psychometric reliability tests.

Discussing *the myth of the metals* raises concerns about the way we frame the *use* of developmental metrics and models. I suggest that we are wrongheaded—ethically and scientifically—when we think that the proper use of developmental assessments is to find out how good people are so that we can position them appropriately in the social group and thus give them the acclaim, trust, and responsibility they deserve. If we take methodological consideration seriously and feel the force of the Basic Moral Imperative **we simply can't use developmental assessments to engineer meritocracies.** To ground this discussion I examine an approach that has jettisoned this myth because it is rigorously separating factual claims from evaluative ones, avoiding holistic vague assessments, and wedding all its efforts to educational interventions.

When all is said and done **I issue a call for a higher level of discourse in the integral Community about the way we use developmental metrics and models.** The vision of a future informed by *rigorous* developmental assessments needs to be realistically and responsibly articulated and re-articulated. **We need concerted philosophical, ethical, and political reflections on role and future of developmental assessments in our society.**

Stein, Z. (2008) Myth busting and metric making: Refashioning the discourse about development. Excursus of Integral Leadership Review. *Integral Leadership Review*. Vol 8. No. 5.

¹ This paper was composed with invaluable help from Dr. Theo Dawson. She is an indefatigable source of support and insight. And she made the figures presented below. I would also like to thank Katie Heikkinen and Clint Fuhs for their editorial comments.

Introduction: toward quality control parameters for psychological technologies

This paper supplements a longer more academic piece that is to be published early next year with *Integral Review* (and follows a less academic preamble published on Wilber's blog and with Integral Life). These publications share a common goal. I'm looking to begin a difficult conversation with those who have an interest in the use of developmental assessment systems. Frankly, I'm concerned about what passes as valid in the current discourse about development and the practices surrounding it. As a philosopher and cognitive scientist pursuing a doctorate in human development and as the senior analyst for a testing service that wields developmental metrics and models, I'm well positioned to offer some constructive criticism.

In essence, I am arguing that we need to come to hold more sophisticated views about human development, our assessments of it, and what we ought to be doing with these assessments in practice. Specifically, I maintain that two myths should be prominently and permanently busted by the Integral Community. *The myth of the given* has been named and is already generally disparaged. The other myth remains nameless but plagues our efforts. *The myth of the metals* is introduced here as the second myth requiring critical attention. Importantly, if we choose to jettison both myths we must refashion the practice of developmental assessment and the discourse surrounding it. The first myth raises epistemological issues and its critique should lead us to pay more attention to how our developmental metrics are made. The second raises social, ethical, and political issues and its critique should lead us to pay more attention to how our developmental metrics are used.²

Below I take up each myth, quickly introduce it, briefly bust it, explain how it applies to our discourse about development, and then give some examples of an approach that jettisoned these myths long ago. For those interested in the details, look to the forthcoming *Integral Review* paper where I cite my sources and justify my contentions more carefully. For those interested in less detail, glance at the figures below to get a sense of the power of an approach that is not wedded to the myths. But before I cut to the chase I want to frame the issues a bit, mainly because even *constructive* criticisms can be hard to hear.

I am afraid that much of what I have to say may fall upon deaf ears because the criticisms will be taken *ad hominem* when in fact they just feel that way. It was the great American philosopher, C.S. Peirce, who first recognized that ideas arise out of the life of communities and thus that the criticism of an idea is in effect the criticism of a way of life. This sentiment was echoed by another great American philosopher, Wilfrid Sellars, who suggested that one task of philosophy is to model the use of sophisticated conceptual frameworks in order cast *shame* upon the language games currently in play. No doubt, I will cast shame upon some approaches and those who endorse such approaches may feel attacked, both ideologically and economically. It is important to recognize that this whole discussion is complicated by the fact that people make money using developmental assessments (no one is getting rich yet, but livelihoods are on the line). And to put all my cards on the table, the approach I endorse is for sale too. But my efforts at exercising quality control in the field are not primarily economical.

² Importantly, what follows is not a criticism of Wilber's psychological model. I offer very general concerns that should be cause for pause regardless of which model holds our allegiance. Of course, Wilber looms large when we are spilling ink about Integral issues. If there is a criticism to level at Wilber in this context it is that his writing is *deceptively clear*. If you read all his footnotes and catch all the caveats you'll see that he's not that far from where I am. But because he writes so clearly, we often follow the story line without digging into the details and think we get him when we don't. The problem is that his writing allows for downward assimilations—oversimplified understandings—and we run with it without grasping its full complexity. The preamble to this piece I posted on Wilber's blog and Integral Life touches on this. That said, this is not the place to demonstrate the kind of sophisticated exegeses Wilber deserves (and rarely receives). This is also not the place to outline the specific criticisms I do have of his psychological model, which stem almost entirely from his relative neglect of Vygotsky and comrades.

As noted, Peirce and Sellars offered insights into the lifeblood of ideas. They also took time to expose the ethical presuppositions of discourses about truth. That is, they were convinced that scientific practices entailed a set of specific ethical commitments, e.g. cooperation, transparency, selflessness, etc. Without sharing these ethical commitments we lose the mutual regard and trust needed to open ourselves to the "peculiarly unforced force of the better argument." Competition in the marketplace has its purpose, but in the marketplace of *ideas* we need to preserve the common causes of Truth and Goodness that keep us honest and reasonable. My concerns are about what we are doing as a community when we talk about and use developmental models and metrics. One of my key points is that we must begin to recognize more indexes of quality in the market place of ideas than the bottom line. *The most popular approaches are not necessarily the most preferable.* The conversation I'm trying to start is an attempt to raise the level of discourse about which developmental approaches are to be preferred, for which purposes, and why.

We need to have this conversation now because we stand on the brink of a new era in which the use of—and thus market for—psychological technologies will be growing rapidly. By *psychological technologies* I refer to a wide array of engineered techniques, measures, and devices that gear into psychological phenomena in a quasi-instrumental manner. The classic examples are the meditative and contemplative practices refined by religious traditions. These were devised and disseminated explicitly as techniques for altering the nature of experience and the self. Contemporary psychological technologies hit upon a wider array of techniques, but the basic idea is the same. From cognitive behavioral therapy to Myers-Briggs and from Big Mind Process to Scientology, we are self-consciously and explicitly utilizing techniques that were made and marketed as purposeful targeted interventions into our psychological lives. Importantly, while the first wave of psychological technologies was developed in the context of cultural movements, contemporary psychological technologies are being developed in the context of market cultures. And as life conditions become increasingly complex the market for psychological technologies will grow.

In the early 1970's the sociologist and cultural critic Daniel Bell wrote a book titled, ***The Coming of Post-industrial Society***. Across the Atlantic at the same time Habermas offered a slim volume titled, ***Legitimation Crisis***. Both books prophetically outlined emerging trends in post-industrial economic and state bureaucracies and in post-modern cultural forms and norms. Bell and Habermas both suggested that part and parcel of the major socio-cultural shift underway are drastic increases in the task-demands of work and life, which require the emergence of a new group of knowledge workers specifically concerned with the psychological complexities of the new socio-cultural milieu. Bell in particular foretold of a shift in the value of different types of knowledge. As techno-economic innovations accelerate we will come to devalue specific types of technical knowledge regarding production processes because this knowledge is destined for obsolescence. Instead, we will begin to value psychological knowledge about the motivational and cognitive processes of the people who face the ever-shifting demands of post-industrial society. Habermas saw a comparable trend affecting political systems. The problems facing state bureaucracies are quickly becoming so incomprehensibly complex that leaders may feel forced to manufacture consent, e.g. via the strategic deployment of psychologically sophisticated advertising regimes. Fueled by this insight Habermas would write prolifically on the psychological demands of a post-modern *democratic* public-sphere and on the need for sociologists and psychologists to arrange ameliorative interventions. The bottom line in both accounts is that there are world-historical structural transformations underway that demand the creation and dissemination of psychological technologies. Psychological technologies are becoming a valuable resource.

Because of this kind of emerging demand for psychological technologies we need to have serious conversations about quality control that look beyond market values. It's already easy to market and sell self-help approaches and consulting practices without research findings concerning their validity or effectiveness. Moreover, the striking popularity of faddish "brain based" educational practices highlights the seductive allure of approaches that claim scientific respectability. The question we should all be concerned about is how we can begin to sort the wheat from the chaff in this climate of increasing demand for psychological technologies. I fear that if market mechanisms prevail we will face a kind of commodity fetishism wherein the value of what we offer is systematically distorted. Recent disturbing trends in biotech and healthcare already show what can happen when market forces mediate the interface of science and the lifeworld.

It is with this context in mind that I'm out to do some myth busting. Developmental models and metrics are a particularly powerful and prominent type of psychological technology, especially for Integral Folks. The discussions I'm looking to start are about how to look past the simple surfaces of these tools and begin asking more sophisticated questions about how they should be made and used. To borrow a line from my colleague and mentor, Howard Gardner, we need to begin a rigorous discussion about how to detect the "symptoms of quality" in the developmental approaches before us. What follows is only a start.

The myth of the given: from knowledge as found to knowledge as made

Wilfrid Sellars quietly ushered in a sea change in academic philosophy when he published "Empiricism and the Philosophy of Mind" in 1956. Along side W. V. O. Quine's "Two Dogmas of Empiricism" and Wittgenstein's *Philosophical Investigations*, this revolutionary piece dislodged the prevailing hegemony of logical positivism and opened up a new era of post-analytical philosophy, an era that would welcome a plurality of approaches to broad questions by accepting the likes of Taylor, Davidson, Rorty, Searle, and Chalmers into the mainstream. And it was in this piece that Sellars originated his startlingly powerful critique of the "myth of the given." This phrase, which Sellars coined, is really a contemporary gloss on a theme in epistemology that can be traced from Sellars through Peirce and back to Kant (from Pittsburg through Cambridge to Königsberg). This is a myth about what we know and how we know it. Jettisoning this myth is an act of epistemic humility wherein we admit that the world appears only in light of the constructs we employ to question it. Knowledge is not simply given via sundry experiences (that's the myth); knowledge is made via experiences that are systematically disclosed. Therefore concerns about how we make knowledge take center stage. This is the heart of the post-metaphysical turn in philosophy.

The busting of this myth has had a big effect on many streams of discourse, especially in the philosophy of science. Generally, as a result we have shifted toward a focus on *representational devices*, which are the things we do and use in order to see what we are interested in. Roughly speaking, scientists are in the business of collectively building, testing, and refining these world-disclosing devices. Thus scientists are *not* in the business of fact finding. If we jettison the myth of the given concerns about *methods* take primacy. We need to be sure our representational devices are well calibrated and reasonable before we start talking about what we have "found." And when we do talk about findings and start telling stories about "what's going on out there," we should issue caveats and remain open to the continual revision of our key methods.

Of course, the myth has persisted in many disciplines, including developmental psychology. The tables in the back of Wilber's *Integral Psychology* offer a compendium of developmental models, displaying the names and descriptions of levels in the unfolding of key competencies. *Models* are about "what's going on out there" (or when it come to human development, "what's going on in there"). Models are a key component in any developmental approach; when we

think developmentally we are thinking in terms of some working model or map of how development unfolds. If we take the myth of the given seriously we should become very concerned about how these models are built.

Briefly, model building in developmental psychology entails the use of various representational devices. Before we build models we need to refine the techniques that allow us to see development, to measure it, and to tease apart the properties of developmental sequences. Building a developmental model requires a metric and a method. So the best way to see if the myth of the given plagues an approach is to look at the *methods sections* of the publications that present the model.

A meta-analysis of the literature surrounding the most popular developmental models leaves us disappointed with their methodological rigor (see my forthcoming paper in *Integral Review* for details). Generally, rich descriptive models and explanations of how development unfolds are produced and consumed (especially in the Integral Community) with very little attention to the metrics and methods that make them possible. We love the knowledge, accept it as true, but could care less about how it was made. That's the myth of the given in a nutshell. If we were to jettison the myth and adopt a properly post-metaphysical approach to developmental theory and practice things would be different.

For example, if we lose the myth then issues surrounding the making and refining of developmental metrics—our key representational devices—would take primacy over the presentation of various stage models and narratives. We would turn away from the stories describing development and towards the making of the metrics that justify these stories. There are a variety of methods for studying and refining metrics—metrology is a whole field, with a sub-discipline called psychometrics. The most basic studies that focus on the validity of developmental metrics are inter-rater reliability studies, which aim at determining the inter-subjective agreement adhering to a metric, i.e. *do you see what I see?* We can also conduct empirical cross-metric comparisons between different assessment systems, which aim at determining the relations and potential redundancy of different assessments, i.e. *do you measure what I measure?* Finally, we can mathematically model the psychometric properties of an assessment system by analyzing the results of empirical studies, i.e. *how consistent is our metric across conditions?*

These three basic types of studies (and there are more) are designed to focus on the properties of an assessment system that are indicative of its validity and integrity. Does the assessment garner inter-subjective agreement? How does the assessment relate to other assessments? Does the assessment change when it is applied in different conditions? Frankly, and to the point, if we are looking to get beyond the myth of the given, then having empirically grounded answers to these questions is a precondition for building a model in developmental psychology. Think about it. How much should we buy into a model that was built with findings that are disclosed unreliably (e.g. the metric yields poor inter-rater agreement; or there are major accuracy differences across assessment conditions)? How much do we understand a model if we don't know how the findings it is based on compare to the findings used to build other models? Importantly, these basic studies can be combined to yield more complex studies, e.g. cross-metric comparisons of inter-rater reliability, etc. Moreover, beyond these kinds of studies we need studies about the effects of different developmental interventions. We need to ask not just "is our metric and model valid?" but also, "is our approach having the effect we think it should?" The key point here is that we have ways to study the worth of our methods and these allow us to couch our models more reasonably and responsibly.

To return to the themes broached above in the introduction, this kind of intra-disciplinary self-reflection represents a level of methodological sophistication that characterizes many sciences,

but it is especially important for those aiming to generate *useable knowledge*. The truth is that as a discipline we have not attained this level of sophistication and, at this point, there is no generally accepted endeavor aimed at exercising quality control vis-à-vis the proliferation of developmental approaches. The FDA is a useful example (both in its failures and successes) of a quality control agency that mediates between researchers and the lifeworld by bringing attention to the methods and claims being made in the laboratory. Were we to thoroughly dislodge the myth of the given in developmental studies and choose to adopt an epistemologically responsible approach to building and using developmental metrics and models we could learn from agencies like the FDA. Not any theory and its concomitant interventions should reach the market. We need to be concerned *first* about what justifies the approach and then about how exciting or revolutionary it is. Right now we are mainly concerned with the latter, which is putting the cart in front of the horse.

Beyond the myth of the given: monitoring methods; monitoring efficacy

In this section I'm going to offer reconnaissance from the field in the hopes of demonstrating the power of a developmental approach that has jettisoned the myth of the given. The Developmental Testing Service (DTS) employs a set of methods for building and using developmental assessments, including a domain general developmental metric known as the Lectical™ Assessment System (LAS). We also wield a sophisticated theoretical framework called Dynamic Skill Theory, which consists of a set of models about human developmental processes. A great deal of information is available on-line about the various facets of our approach. The LAS and its robust analytical accouterments were developed by Dr. Theo Dawson, founder, president, and CEO of DTS. Information is available online at www.devttestservice.com or www.lectica.info. Dynamic Skill Theory, which set many of the empirical and theoretical foundations for the creation of the LAS, is the result of over three decades of research by Dr. Kurt Fischer, Charles Bigelow Professor of Cognition and Education at Harvard and founder of the International Mind, Brain, and Education Society. Numerous empirical studies using and refining this theory can be found here: <http://gseweb.harvard.edu/~ddl/>. For an overview of the approach especially articulated for Integral Scholars see the paper I co-authored with Katie Heikkinen that was recently published in the *Journal of Integral Theory and Practice* (a version of this paper is available on the DTS website).

Given the material available on-line, this is not the place to outline the details of this complex approach. What I want to do is highlight some of its features that bear on the issues at hand. I believe this approach sets the bar very high in terms of the methodological sophistication needed to overcome the myth of the given. I'm not going to focus on the unparalleled volume and quality of empirical work supporting the metric and model (although see my forthcoming paper in *Integral Review*). It's easy to establish that we've done more to validate our approach than anyone else by just comparing publication records.

Instead of harping on that string, I will focus on the way we think about our methods. We have devised a broad approach for research and practice that systematically integrates procedures for continually testing and refining our methods. That is, we are not just running with what we have; instead, we are perennially concerned about the worth of what we are doing. So with every new project we retest our instruments to make sure they are working the way they should. We also do all we can to monitor the effectiveness of our endeavors by looking into what happens as a result of our interventions. Importantly, both the monitoring of methods and the monitoring of efficacy are empirical affairs. And we make our findings about our own practice available, either via peer reviewed journal publications or as reports on the DTS website.

We monitor the efficacy of our interventions by establishing a very specific type of collaborative relationship with our clients—an approach we call *developmental maieutics* (see figure 1). Generally, this approach is about how to collaboratively and responsibly build usable knowledge. For example, a number of years ago we were approached by a network of major government agencies looking for help with a broad leadership development initiative.³ Over a series of research studies and educational interventions we unfolded an endeavor roughly aligned with the steps outlined in figure 1.

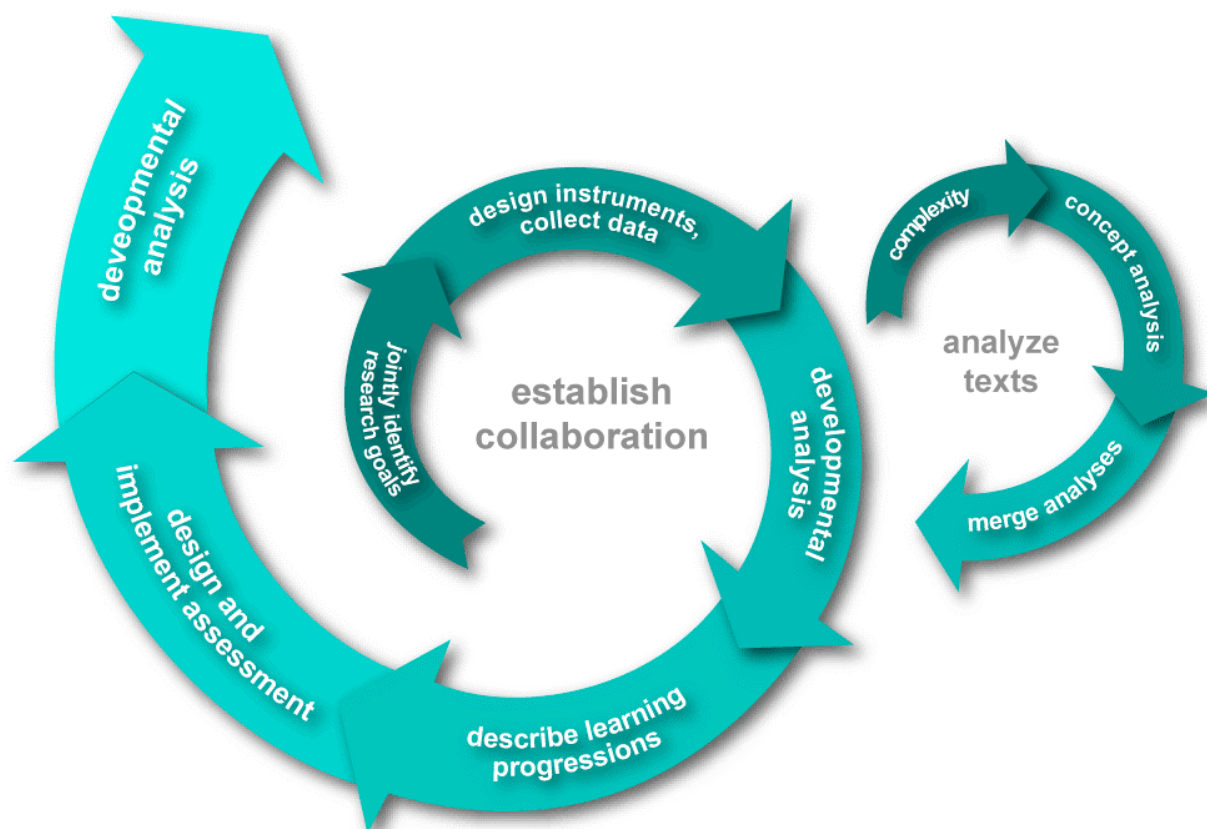


Figure 1: The *developmental maieutics* spiral is a representation of how we systematically combine research and practice and monitor the efficacy of our endeavors.

First, we co-constructed a set of key research questions geared into their needs, which we operationalized in terms of specially designed formative assessments. These assessments yielded data that were submitted to developmental analyses resulting in unique and powerful rational reconstructions of the leadership domain. The insights we gained about leadership development were then used to frame the reformation of professional development programs. These reforms were researched and the efficacy of our developmentally sophisticated leadership curriculum was determined (we found it worked better than traditional curricula). Then we retooled and launched another iteration of formative assessments that have been

³ Incidentally, these people knew Integral and they went shopping for developmental approaches. They did their homework and looked at the benefits and liabilities of various methods and models before they came to us: a good example of the critical consumption of psychological technologies.

scaled up for general use by organizational consultants and educators. This latest round of assessments will yield new information about the development of leadership capabilities, which will inform educational interventions on our part, the effectiveness of which we will monitor. And so on it goes as we monitor the efficacy of our efforts in order to continually improve our practices (all the relevant reports are available at www.devtestservice.com).

Importantly, all along we were also monitoring the validity and reliability of our methods (see figure 2 for an overview of what we mean by *validity*). Our approach allows us to assess development in many different lines or domains, so we are able to generate a variety of psychographs using a single methodological approach (see figure 3). But whenever we enter a new domain (or are dealing with unique properties of a domain we've already researched) we recalibrate aspects of our metric. We also recalibrate when we are doing assessments in a new context, such as online leadership assessment vs. interview based leadership assessment. This recalibration has several facets, including determining the new levels of inter-rater reliability and mathematically modeling the performance of the metric in the new condition. So we are continually empirically testing the accuracy of our data gathering instruments, which are, importantly, the very assessments we use to determine the level of performances of persons in our studies (the ethical dimension of this continual concern about accuracy and objectivity will be unpacked in the next section).

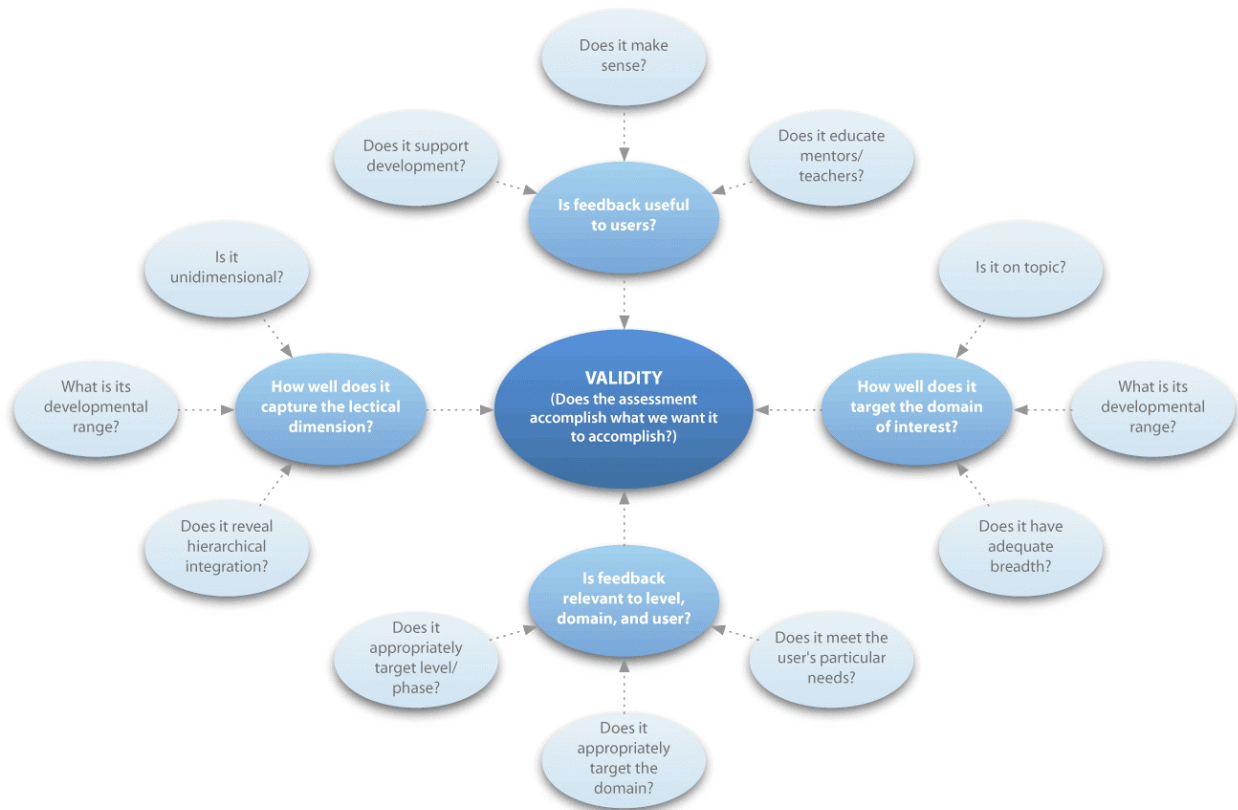


Figure 2: Displays a concept map about the complexity of making a valid assessment. This figure is specifically about how we consider the validity of our Lectual assessments, which are geared into the hierarchical unfolding of skills and concepts in specific domains (i.e., lines). Questions about feedback modalities (top & bottom) are as important questions about the accuracy (left & right).

From where I sit, this degree of concern about methodological issues is not optional; it represents the kind epistemological responsibility we need to assume if we are looking to jettison the myth of the given. For this paper I've put aside issues about who is publishing where (or at all) on these kinds of issues and which approaches have opened themselves to rigorous peer-review (although see my piece forthcoming in *Integral Review*). Here I am just focusing on the way we *think* about the methods we use. And I'm concerned both about the thinking of experts in the field who employ developmental approaches and about the thinking of those who consume these approaches.

The way things stand now we are buying and selling this stuff as if it is something that it is not. We are building businesses and markets around a product the true dimensions of which we often appear to systemically misunderstand. Developmental models and metrics are *not* fixed and final constructs built on a foundation of clear incontestable facts. They are *not* culture and context transcending characterizations of human personalities nor do they offer indisputable and definitive insight into lives of individuals. If we are making these kinds of claims (implicitly or explicitly) we are trafficking in myths, especially if we are not willing to put our methods where our mouth is. Overcoming the myth of the given means admitting the provisional, bounded, and multi-perspectival nature of all models and metrics. It also entails, as exemplified by the work underway at DTS, adopting procedures that can guide a recursive process for continually refining research, theory, and practice. We are a long way from a field where this level quality control is the norm. The myth of the given is to blame. But this myth draws its life from another.

The myth of the metals: psychometrics and meritocracy

The second myth is Plato's. In the *Republic* he suggests that philosophically inclined social engineers should devise a myth in order to justify the structure of a society that is organized hierarchically according to capabilities and dispositions. The *myth of the metals* goes like this: some citizens have blood mixed with Gold, others Silver, Iron, or Copper. This metallic endowment defines a person's essence and allows their being ranked and assigned a role in society. Assessments of traits and capabilities serve this differential distribution: Gold and Silver are indicative of Leaders and Warriors, Iron and Copper of Merchants and Farmers.

The myth of the metals is way of framing the use of psychological assessments. It suggests that such assessments are capable of defining the essence of a person and determining the range of what is possible and preferable for them. Plato is concerned with justice and believes that the contingencies of human nature make it necessary to engineer a harmonious society. He envisions a complex and radical public educational system, with various forms of psychological and physiological assessment for evaluating individuals and putting them in their place. The myth of the metals is essential to this task. The idea is that if we plan to use assessments of capabilities in the structuring of society then we must create ideologies to justify the differential distribution of opportunities that result from those assessments. Such an ideology is to be devised and disseminated by the leaders as a way of making sense of and enforcing a *caste system* to those incapable of grasping the abstract ideal of justice, which is its true justification.

Now, it was Karl Popper, who in 1945 published *The Open Society and Its Enemies*, and first made absolutely clear the totalitarian bent of Plato's political vision. Moreover, he argues that the myth of the metals is the lynchpin of the system because it masks coercive social engineering practices by disguising them as the fateful and acceptable decrees of authorities. But for the sake of argument let us step back from alarmist accusations of totalitarianism (Popper was writing in the shadow of the Nazis), give Plato the benefit of the doubt, and read his political system as a kind of *meritocracy*. This weakens the rhetoric and allows us to frame the myth of the metals in less controversial (but still not unproblematic) terms. The myth of the

metals appears these days as a set of ideas that frame the use of psychological assessments for social purposes, suggesting that they provide us enough insight into the essence of people that they would allow us to engineer a meritocracy.

It is interesting to note that this very issue is raised in the books by Bell and Habermas that set up our discussion about the burgeoning markets for psychological technologies. Right along with their predictions about the shape of post-industrial and post-modern socio-cultural trends are predications about emerging modes of social control and organization, which they see as increasingly tied to the deployment of psychological technologies. Both suggest that the increasing complexity of life conditions will require that bureaucracies facilitate the maintenance of social-role performance through the continual assessment of capability and motivation. But they have different levels of optimism about the effects of these efforts. Bell sees these trends as bringing about a kind of meritocracy that supplants outdated barriers and biases, replacing them with scientific indexes of excellence. Habermas is less optimistic; in fact he is worried. He argues that scientific mechanisms for social stratification are liable to misuse because they risk being engineered and reified ideologically, thus suggesting the possible emergence of a *technocracy* run by elite social engineers. This is a possibility he worked to counteract by demonstrating over the course of his career that methods of human resource management conducive to the maintenance of techno-economic systems may not be likewise conducive to the maintenance of democratic forms of life. The traits we assess and deem worthy of promotion hinge upon the values we are looking to promote.

In any case, while the ideal of justice as a harmonious system of rights and responsibilities stands—despite the flaws of any particular vision regarding its institutionalization—there are serious problems with the notion of handing someone his or her identity and role in society based on a set of assessments administered by a small group of experts. A belief in omniscient and omnipotent assessments that could be used to engineer a meritocracy is fundamentally wrongheaded, on both methodological and ethical grounds. But something like this—something like the myth of the metals—seems to be evident in many cultures that employ (or are looking to employ) developmental assessments. That is, there seems to be a belief that developmental metrics and models have the power to characterize the essence of a person and that the primary use of these characterizations should be for social-role identification.

This myth is evident in the fact that many who consider themselves developmentalists think it is possible to make holistic developmental assessments that determine who a person is. It is also evident in the hypertrophy of higher-levels, which results in the belief that *more developed people are better people*. Generally, when we buy into the myth of the metals, it is thought that we use developmental assessments to find out how good people are so that we can then give them the acclaim, trust, and responsibility they deserve. I may be overstating my case here, but I think that even a cursory familiarity with the current discourse surrounding the practice of developmental assessment reveals that it looks like a lot like the myth of the metals.

Busting this myth means refashioning the discourse surrounding the use of developmental assessments. This task has at least three facets. There are two basic points about the limits of developmental theorizing and measurement. And there is one important ethical point about how developmental metrics and models ought to be used.

The first basic methodological point concerns the difference between *objective descriptions* and the *evaluative prescriptions*. Since Kohlberg first saw it, the *naturalistic fallacy* has plagued developmental psychology. This classic philosophical issue can be traced to Hume and Kant, who argued convincingly about the error of attempting to derive values from facts. Kant in particular was at pains to demonstrate that merely factual accounts about the genesis and structure of cognitive processes tell us nothing about their worth or validity. This same kind of

argument reappeared roughly a century later in the writings of Gottlob Frege and C.S. Peirce, who offered devastating critiques of psychologism at the dawn of experimental psychology. The basic idea is that psychology is a *fact-stating discourse* offering descriptions and explanations, and yet many psychological phenomena are open to being the topic of *evaluative discourses*, above and beyond their being merely described. The statement that you have an IQ of 100 does not come with a value attached. We have to tag facts with values, e.g., by deciding that an IQ of 100 is too low for Mensa membership. This means that a separate discourse needs to take place, one determining the value of specific psychological facts.

But it appears that we are easily seduced into bootstrapping our languages of evaluation from the languages we use to objectively describe developmental patterns and pathways. The default position in the discourse now is that *higher is better*. However, centuries of philosophical hand wringing about the naturalistic fallacy should teach us that determining the value of being at a level is different from determining the objective fact that one is at a level. Simple growth-to-goodness models overlook the radically non-obvious evaluative import of being assigned a level score. Facts and values are not the same, although the myth of the metals would have them be.

The second basic methodological point ties back into the first myth. There is overwhelming evidence showing that our metrics are limited and that we can't touch the true complexity of human development. In this light, the idea that a holistic assessment could tell us about the *essence* of a person is absurd and flagrantly ideological. Developmental assessments at their best can only paint pictures of the differential distribution of capabilities within persons. We can't assess people as a whole, we can only assess their performances along particular lines in particular contexts. And performances vary across contexts, which means that you may perform at one level in one context and at a very different level in another context.

Intra-personal developmental variability is ubiquitous and throws into doubt the validity of blanket generalizations about *who a person is* or what they are capable of based on the results of even the best assessment. That is, if we take all methodological caveats into account, it is fundamentally wrong to think of a person as being *at a level*. Individuals perform at different levels along different lines and at different levels along the same line in different contexts. We are all over the place, and no developmental assessment will ever capture our true complexity. Yet the myth of the metals would bestow on some group of experts a unique kind of knowledge and insight that vastly outstrips the kind of knowledge gained through the responsible use of a developmental assessment system.

Finally, methodological considerations aside, there is an important ethical point here about the use we make of our evaluations of others. A century before Michael Foucault, Emerson, the sage of Concord, drew up the equation of *knowledge and power*. He wrote in his essays that every new fact is also a new weapon in the arsenal of power, that every move toward the acquisition of knowledge is simultaneously a move to tap sources of influence and dominance. Now, Emerson was looking for knowledge with the power to liberate, but he saw that he would be wielding a double-edged sword. These concerns about the use we make of the knowledge we have led Emerson to assert the primacy of character over intellect. Knowledge is neutral regarding its use: we can use medical knowledge to torture or to cure. Thus we must frame the use of knowledge and guide its acquisition in light of explicit and articulate ethical commitments (e.g. the Hippocratic oath).

This is a very general issue about the role of psychological technologies in culture and society. Habermas's concerns about the technocratic deployment of techniques for human resource management are to the point. We should be very concerned about the basic ethical frameworks that guide our use of developmental assessments. Importantly, in my view this comes down to the classic Kantian ethical decision: do we relate to others as ends-in-themselves, or as means-

to-an-end? We can also read it in light of Wilber's Basic Moral Imperative: are we promoting the most depth for the most span? According to both views it is clear that we should administer developmental assessments in order to promote development, not just to rank people and assign them their position in an organization or social group. The myth of the metals would have us use assessments solely to administer the allocation of human capabilities and to inform us about the relative worth of one another. If we lose the myth then we must rethink this use. It seems vastly preferable to wed assessments to educative efforts at all levels and use assessments solely for the purpose of promoting development.

Beyond the Myth of the Metals: democracy and education (and psychometrics)

Here, once again, I will offer an overview of the approach we take at DTS in order to show what it looks like to lose the myth. Our approach avoids the kinds of methodological and ethical problems that plague many developmental approaches. Once again, numerous publications and reports concerning these issues can be found on our web site. It may be worth noting, as a way of framing what follows, that the unpopularity of developmental approaches in the academy can be traced to this myth; we are sitting ducks for post-modernists if we don't couch our models and metrics more carefully. Simple growth-to-goodness models oversimplify evaluative issues and fly in the face of hard won bids for increased pluralism concerning worth, identity, and rationality. Illusions about the omniscience of our assessments appear naive in light of post-analytical philosophy of science, not to mention how our claims to measure the *essence* of a person conflict with the value post-modernists place on the singularity of the individual—who remains always beyond the reach of our objectifying gaze. Finally, the use of developmental assessments for the assignment of social roles is hard to swallow for those who are sensitized to the injustices and liabilities of even the most well-meaning modes of social control.

To start, our methods separate questions of fact from questions of value. Following in the wake of Habermas's work on methods in the social sciences, and his constructive criticisms of Kohlbergian approaches to moral development, we promote a division of labor between psychology and philosophy. This kind of approach is facilitated by the nature of our assessment system, the LAS, which rigorously separates the *content* of performances from their deeper *structure*. Just how this works is a long story (see Stein and Heikkinen in the *Journal of Integral Theory and Practice*). The big take home is that there are many types of performances that can occur at the same level. It is the job of psychology to describe these performances and assign them a level score. It is the job of philosophy to determine the relative worth of the different performances. These are two distinct discourses that need to take place concerning any given performance: psychology tells us *where it is at*; philosophy tells us *what it is worth*.

Thus, if we compare two performances and determine—using our assessment system—that one is at a higher level than the other, we cannot simply assume that the higher one is to be preferred. We must ask a whole host of other questions about the performances, questions about their coherence, appropriateness, success, etc., i.e., questions that bear on the value of the performances. Of course, this requires that we clarify our commitments (both epistemological and ethical) about what makes for a good performance. This entails a thoroughgoing familiarity with the domain in question. For example, evaluative issues in the domain of leadership are tremendously complex. We worked for years with leadership experts and engaged countless texts before we could include evaluative feedback into our assessments. That is, we could measure leadership reasoning (i.e., assign it a level score) long before we could evaluate it (i.e., give feedback to condemn, praise, or prescribe). And we remain continually sensitive to how these central evaluative issues shift and change in different situations. Our methods land us far from notions about development as a simple process of growth-to-goodness.

Our methods also land us far from holistic claims about the essences of people. We assess performances not people. And performances take place in specific domains, at particular times, in particular contexts. Figure 3 displays a diachronic psychograph focused on leadership, which displays how the differential distribution of leadership capabilities within a single person changes over time. This is the kind of assessment modality that we devised during our work with the government. Importantly, developmental change in the domain of leadership is not linear, and changes in different lines take place at different rates. This is how we think about development in any domain. A single score can never summarize an individual. Thus we work to construct dynamic developmental profiles.

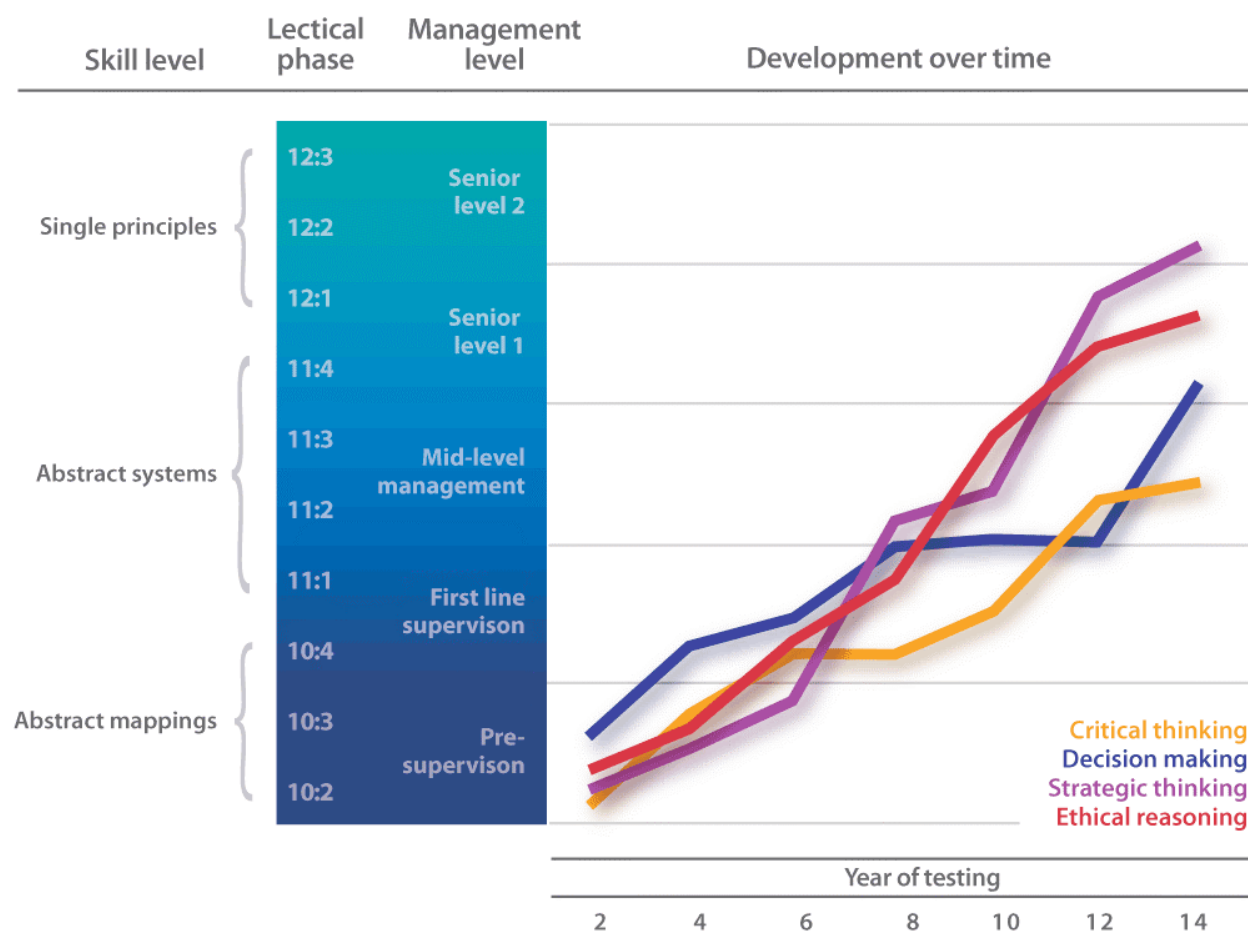


Figure 3: Diachronic psychograph focused on leadership, i.e., this psychograph traces the development of several lines within the leadership domain over time. This is the kind of feedback that flouts the myth of the metals. For information on the levels and phases go to www.lectica.info, or see Fischer's levels in the tables in Wilber's *Integral Psychology*, where our levels are lined up with levels in other systems.

A second issue also looms large when we think about the dynamic development of individuals. Figure 4 is a way of displaying how contexts affect the level of performances. A useful way to think about this is in terms of two key developmental constructs that Fischer has researched for decades: functional-level vs. optimal-level. That is, in any given domain individuals' performances are best thought of in terms of a *developmental range*. We perform at a higher level—our optimal level—in supportive and familiar contexts, whereas in challenging or unfamiliar contexts we drop down to our functional level. The difference between functional and optimal levels can be very large (spanning as much as 3 or 4 levels). Throw in issues of stress, emotion, and interpersonal dynamics and variability of level across contexts becomes a major issue in the study of development. As I said above: we perform at different levels along different lines and at different levels along the same line in different contexts.

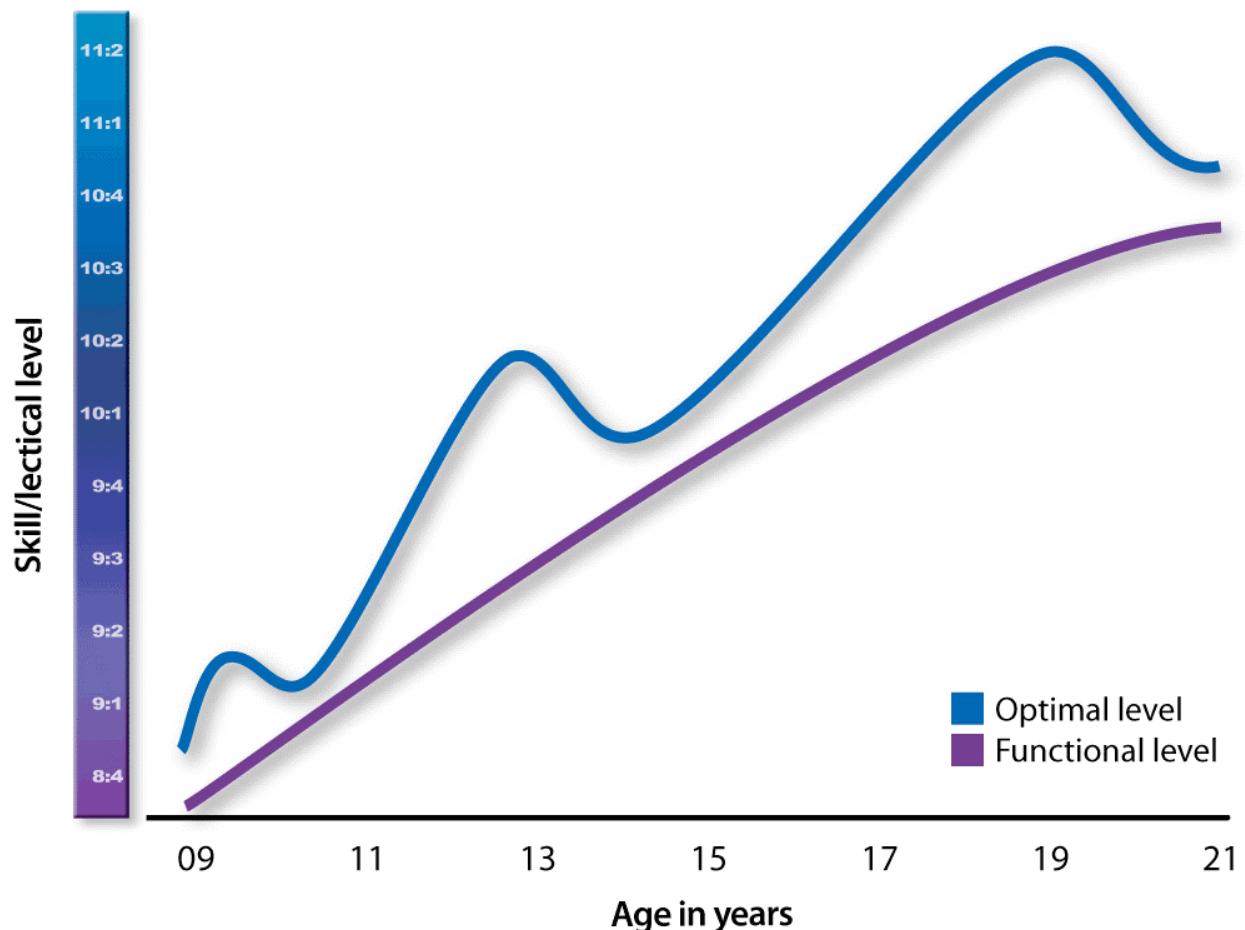


Figure 4: This figure displays changes in functional and optimal levels over time. Functional-level performances are those that occur without support in challenging contexts. Optimal-level performances are those that occur with support or in familiar contexts. As the figure shows, at any give time functional-level performances lag behind optimal-level performances as much as a level or more. The difference between functional and optimal levels at any given time is an individual's *developmental range* (i.e. above the *developmental range* is greater at age 13 than it is at age 15). This pattern of variability in performances has been demonstrated empirically in variety of domains by Fischer and was first researched by the great Russian developmental psychologist Lev Vygotsky. If we want to jettison the myth of the metals it is crucial to account for this kind of variability when administering and interpreting developmental assessments.

Given this variability, both across domains and lines and across contexts, we must admit that even the best assessment can give us nothing more than a passing snap-shot of an individual. So there are good methodological reasons for discouraging certain uses of developmental assessments. The complexity of who a person is and what they are capable of will always remain beyond the grasp of our assessments. If we are humble about what our assessments can accomplish then we must be careful to frame their use responsibly. Their use as instruments for social role allocation (e.g., hiring, firing, or promotion) is simply wrongheaded. The only index of how a person will do on the job is how they have done on comparable jobs. The stakes are too high and our measurement instruments are too crude to use the results of a single assessment to determine the future of an individual or an organization.

At DTS we are careful to insure that our assessments are used mainly for educative purposes. That is, while we would never support their being used for simple social role allocation, we do think that a single assessment can tentatively determine the profile of an individual's competences with enough accuracy to warrant their being assigned to specific types of educational interventions. Developmentally appropriate pedagogical interventions can range from assigning books at a particular level about a particular topic to placement in a focused professional development program. But an individual's responsiveness to these interventions should be monitored via further assessments. Assessments should never be used as fixed and final indexes about a person, they should be used as ongoing sources of information to frame that person's further development. Doctors don't take your temperature once and then judge your overall health. They take it multiple times over the course of treatment and adjust interventions accordingly. At DTS we administer developmental assessments in order to promote development.

This is a simple and seemingly uncontroversial idea. But it flies in the face of the myth of the metals, where development is measured in order to determine how good someone is and to bestow upon them the acclaim and responsibilities they deserve. Importantly, broader ethical considerations support our ideas about the educative use of developmental assessments and the myth of the metals can be disparaged on these grounds alone. If we follow the discourse about democracy from Jefferson through Dewey and Habermas, it's clear that the conditions that support *just government* look like educational environments that support autonomy. The engineering of a meritocracy does not allow for the reciprocity and fairness required to legitimate democratic regimes because it entails overriding individual autonomy for the sake of collective ends. In our post-modern socio-cultural context the role of developmental assessments—and psychological technologies more generally—should be to foster the autonomy of individuals. From where I sit this means engineering educational environments that have sensitive and accurate assessments embedded in them in order to better facilitate the distribution of educational opportunities that help individuals help themselves.

Conclusion: now what?

We have covered a lot of ground. In light of general concerns about the use and dissemination of psychological technologies I discussed two myths that need to be busted if we want to start exercising quality control in the burgeoning markets for developmental assessment. Considering *the myth of the given* led us to address the nature of our methods, suggesting that we need to systematically study the models and metrics we employ to ensure they are working properly. Considering *the myth of the metals* led us to address how we frame the use of developmental assessment systems, suggesting that we need to separate facts from values, be humble about what we can really measure, and use assessments to help educate people, not merely to rank them.

The arguments I have laid out here will be further advanced in my forthcoming paper for *Integral Review*. My goal here was just to start a conversation not to offer a rigorous and comprehensive treatment of these complex issues. That said, I think there are some important lessons to be drawn at this point, however tentative. It seems clear to me that we need some kind of overarching quality control agency for regulating the markets for psychological technologies. I've been concerned here about developmental assessment systems and their use. But my arguments here could be expanded to implicate a wide variety of other areas. People get hired and fired after taking Myers-Briggs assessments, which boasts disconcertingly low levels of methodological reliability and validity. This kind of misuse of psychological technologies should be flagged and stopped. But we can't do that until we have some visible and reputable agency that is responsible for setting and enforcing quality control parameters.

When it comes specifically to developmental assessments, we need to raise awareness that reliable and accurate developmental assessments are hard to make. We've only been working on them for about 50 years. So it should come as no surprise that the best methods for assessing developmental change in persons *have yet to be invented*. This kind of epistemic humility is scientific and its council is one issued in all applied sciences: work with the tools you have, but be on the look out for better ones. Thus, despite our enthusiasm as to its prospects, the science of developmental assessment is just now building the tools that will usher in an era of mature scientific practice. James Mark Baldwin, who first glimpsed the possibility of a thoroughgoing developmental and integral psychology, was pioneering up through WWII. Piaget died in 1980. His students are still with us today. The aforementioned progenitors of developmental science were prodigious enough to leave room for various progeny. We face not one, but many, developmental psychologies. Time will tell the fate of each. Clearly, as discussed above, questions about their relative *validity* and *utility* bear on questions of their probable *longevity*.

If we begin to openly disparage the myths discussed above we will need to refashion the practice of developmental assessment and the discourse surrounding it. Of course, when we lose one myth we often create another. The vision of a future informed by rigorous and accurate developmental assessments needs to be realistically articulated and re-articulated. We need concerted philosophical, ethical, and political reflections on role and future of developmental assessments in our society. The question: "what should we want from developmental assessments?" is a good one. As are the questions: "what can we realistically expect from them now, in the future, and in principle?" Steering a trajectory forward requires a vision for the future of the discipline that grapples with both what is *possible* and what is *preferable* and an organization of concerned psychologist and practitioners to insure that what is most *preferable* is also most *probable*.