Intuitions of altitude:
Researching the conditions of the possibility for developmental assessment

Zachary Stein
Developmental Testing Service
Harvard Graduate School of Education

Draft for the 1st Biennial Integral Theory Conference, Pleasant Hill, CA. August 2008

> Many of the earliest experiments involving thermometers read like investigations *of* that new instrument not investigations *with* it. How could anything else have been the case during a period when it was totally unclear what the thermometers measured?
>
> -Thomas Kuhn (1977, p. 218)

Introduction: Metrics, Methods, and the Milky Way

In 1875 Charles S. Peirce received funding from the National Academy of Sciences to conduct what would be the first psychological experiments on the North American continent (Cadwllader, 1975; Peirce, 1878; Peirce & Jastrow, 1885). All this research employed statistical techniques developed in Europe for psychophysical experimentation on the liminal properties of sensations. In the most ambitious and original of these experiments, Peirce (1878) used a complicated set of apparatuses to conduct experiments in psychometrics establishing the reliable scalular quality of self-reports about the relative intensity of points of light. However, the ultimate goal of these experiments was to insure the reliability of his eyes for their use in astronomical observations. That is, he was out to clarify the invariant psychophysical structures serving as the conditions of the possibility for measuring the magnitude of stars. With the heart of his method thus justified, Peirce began cataloguing stars by eye in order to support his hypothesis—highly original at the time—that the Milky Way galactic cluster is shaped like a disc.

These were not Pierce's only forays into experimental *metrology* (the science of studying and building metrics). He was, in fact, one of the great metrologists of his generation and a prominent member of several national and international scientific organizations for the standardization of weights and measures (Fisch, 1986). Peirce was actively involved in the

trans-disciplinary and trans-national calibration of a variety of important metrics, including, for example, being primarily responsible for establishing a universal standard for the *meter* based on wavelengths of light in 1878.

I begin with these historical reflections because I do not think it is a coincidence that Peirce was both an eminent metrologist and a key progenitor of current post-metaphysical philosophical approaches (Habermas, 1992; Wilber, 2005). By relinquishing philosophy's claim to a unique privileged mode of insight, Peirce (1868) recast philosophy in terms of the same fallibilistic self-understanding that characterizes scientific endeavors. Understanding philosophy in this way entails turning away from speculative metaphysics and toward the rigorous analysis and critique of inquiry itself. After we abandon the ambitions of *prima philosophia* we are left with the task of positioning philosophy within the system of sciences as "the method of methods" (Peirce, 1882), i.e. the reflective and self-critical operation of the procedural rationality immanent in methodologies. Importantly, and to the point, Peirce was well aware that most successful methods entail the use of good metrics. Therefore, part of the critique of methods (the *sine qua non* of post-metaphysical philosophy) is the critique of metrics. And as Peirce's efforts at measuring the Milky Way make clear, the critique of metrics, like the critique of methods, requires that we bring attention to the conditions of their possibility.

This paper presents the beginnings of experimental work in psychometrics wherein I am attempting to bring attention to the conditions that make it possible for us to *measure minds* (Stein, in prep). As Peirce worked to make clear the invariant properties of his own eyes prior to his astronomical observations, I am working to make clear the invariant properties of our pre-theoretical judgments about the development of thought and language. Ultimately, my reflections on metrics, models, and methods should amount to a kind of prolegomena to any future developmental assessments. That is, I am turning away from the domain of developmental psychology and toward the conditions of its possibility, mainly for the purposes of re-conceiving the domain itself.

This radical move involves tracing the methods, models, and metrics of developmental psychology back to their roots in the pre-theoretical structure of the lifeworld (Habermas, 1988; 1990). Many theorists (e.g. Piaget, 1971; Husserl, 1970) maintain that we can trace physical metrics like rulers and scales—and all the science that depends on them—back to the reliable capabilities of our organism and concomitant sensorimotor practices that attune us to invariance in the world. These basic skills of being-in-the-world congeal into everyday practices that we take for granted (i.e. they form a part of the lifeworld). And it is in the course of working to clarify this background of informal practices that we begin to build intersubjectively calibrated systems of measurement. However, what we can rely on and take for granted when we interact with *things* differs from what is invariant in our interactions with *people* (Baldwin, 1906; Habermas, 1988; Wilber 1995). Efforts at explicating deep-seated social aspects of the lifeworld have been dubbed, *rational reconstructions* (Habermas, 1988; 1990). Thus, as the invariant psychophysical structure of Peirce's eyes accounts for the reliable differential responsiveness that justified their use in measurement, so the "invariant [psycho-social] properties and constitutive rules of the primary lifeworld…can be made explicit [i.e. be rationally reconstructed] as a system of reference [allowing] for the transformation of communicative experience into measured data" (Habermas, 1988).

My broad claim is that we should understand developmental assessments as rational reconstructions of a kind of deep-seated intuitive knowledge that is *always already* a part the network of practices and beliefs that constitute the lifeworld. Just as we unreflective wield an intuitive knowledge of brightness, distance, or weight (based upon certain invariant relations between organism and environment) we also unreflectively wield an intuitive knowledge of development (based upon certain invariant psycho-social structures). Importantly, in both cases, "the expression *intuitive knowledge* should not be understood as meaning that…this knowledge is incapable of being discursively justified. On the contrary, implicit knowledge has to be brought to consciousness…through a well thought-out maieutic method [because] ascertaining the so-

called intuitions is already the first step toward their explication" (Habermas, 1998, p. 40). The research presented here aims to maieuticly explicate the pre-theoretical knowledge underlying developmental assessments. It is motivated by the idea that an "experimental comprehension of the … structure of the lifeworld [is] the precondition of any reliable measurement…" (Habermas, 1988, p. 110; see also Fisher 2004; 2005).

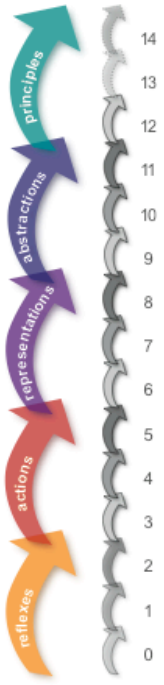
Indeed, the reliability and validity of measurements in developmental psychology have been less than optimal in part because we have not taken the time to clarify the presuppositions of our methods. But there is more to this than improving our assessment techniques. I think that, "by reflecting on the conditions of possible research…sciences articulate their self-understanding" (Habermas 1988, p. 44). To gain a better sense of what the field of developmental psychology is and ought to be we need to bring our attention to some of its most basic features, which means seeing it in light of the primordial perspectives and practices that characterize our humanity more generally. Unlike the methods we use to understand *objects in the world*, we use methods in developmental psychology to understand *ourselves*. Methods based on rational reconstructions are introduced, necessarily, "in connection with categories that must be derivable from the self-understanding of the very subjects [being studied]" (Habermas, 2001, p. 10). This insight is important and ultimately suggests that developmental assessments are simply attempts to improve upon the ways we have always already assessed each other and ourselves. Understanding developmental assessments in this way has the potential to recast ethical concerns about their use in schools, industry, and government. It also has the potential to open up the field toward a wider array of issues, for example, suggesting a general approach to rationally reconstructing levels and skills typically beyond the purview of standard research. I will discus these implications in the conclusion.

Below I begin with by briefly touching on two approaches that suggest a common trait characterizes the majority of developmental assessments. Wilber (2005) and Dawson (2008) both lead me to the idea that a wide variety of developmental metrics are best thought of as

rational reconstructions of some common and basic pre-theoretical knowledge concerning the development of thought and language (i.e. an *intuition of altitude*). I then take a closer look at how to think about rational reconstructions in this context. This sets the stage for the presentation of preliminary results from an empirical study of intuitive judgments about the development of thought and language. This study involves a developmental sort task in which participants are asked to sort a series of statements developmentally and then explain why they sorted them the way they did. Results reveal intriguing trends in both the accuracy and justification of intuitive developmental judgments, both in general and in different age groups.

Different views of the same river: the varieties of developmental assessment

Wilber (1999) brought a great deal of clarity to the field of developmental psychology by demonstrating that, in theory, the vast majority of existing models can be aligned along a common developmental continuum. He offers the construct of *altitude* as an orienting generalization, posting a generic developmental space framing the integration various models of psychological development. Dawson (2001; 2002; 2004) did a similar thing, but instead of aligning models, she aligned metrics. Moreover, instead of offering an orienting generalization she offered empirical evidence suggesting an underlying trait or dimension common across a variety of different developmental assessment systems. Dawson's work was in the service of validating and refining a developmental assessment system that honed in on this underlying dimension. The result was the Letcial™ Assessment System (LAS), which is fruit of over 100 years of research in cognitive developmental psychology (Dawson, 2008; Fischer & Bidell, 2006; Stein & Hiekkinen, 2008). Figure 1 summarizes the results of Dawson's efforts at aligning assessment systems and displays how the levels specified by the LAS line up with the levels specified by a variety of other assessment systems.



	Skill levels (Fischer)	GSSS† (Cmmons)	SISS† (Kohlberg)	GLSS† (Armon)	RJ† (Kitchener & King)	SOI* (Kegan)	SD* (Beck)	Perspective* (Cook-Greuter)
14	principled systems	cross-paradigmatic					coral	cosmic
13	principled mappings	paradigmatic	stage 6		stages 6 & 7		global holistic	global/6th person
12	single principles	meta-systematic	stage 5	stage 5	stage 5	interindividual	systematic/ integrative	5th person
11	abstract systems	systematic	stage 4	stage 4	stage 4	institutional	relativistic	4th person
10	abstract mappings	formal	stage 3	stage 3	stage 3	interpersonal	individualistic achiever	3rd person
9	single abstractions		stage 2				absolutist	
8	representa-tional systems	concrete		stage 2		imperial	power gods	2nd person
7	representa-tional mappings	pimary	stage 1				magical/ animistic	
6	single representations	pre-operational		stage 1		impulsive		1st person
5	sensorimotor systems	sentential						
4	sensorimotor mappings	nominal						
3	single sensori-motor actions	sensory-motor						
2	reflexive systems	circular sensory & motor						
1	reflexive mappings	sensory & motor						
0	single reflexes	calculatory						

*We employed the LAS to determine the hierarchical complexity of the task demands embedded in the level definitions for this sequence. †We have studied the empirical relation between the LAS and the scoring system for this sequence.

Figure 1: Developmental assessment systems aligned along the common dimension operationalized by the LAS

The idea common to both Wilber and Dawson is that the majority of developmental models and metrics are best understood as "different views of the same river" (Wilber, 1999). Sticking with this metaphor, where the river is the rich and inexhaustible complexity of the life-course, it is worth noting that whereas Wilber aligned different accounts of *how the river looks* (i.e. different models and rich descriptions), Dawson aligned different accounts of *how we look at the river* (i.e. different metrics and assessment systems). This is an important difference that bears on the research presented below.

Developmental models are detailed descriptions and explanations of developmental processes. For example, Piaget (1971) and Kohlberg (1969) famously offered a set of models,

consisting of descriptions of various stages and processes to explain stage transitions, etc. Developmental assessment systems (or metrics), on the other hand, are the backbone of the methods that generate and verify developmental models. They consist of criteria for making judgments about the relative development evidenced in specific types of performances (e.g. actions, speech acts, writings, etc.). Importantly, familiarity with a specific model (e.g. knowing how all the levels are described) does not entail that one could accurately assess development in terms of that model. The accurate use of a developmental assessment system takes (sometimes months of) training. Moreover, while specific metrics and models are related they are somewhat independent, in that they can iteratively inform and refine one another. Thus, understanding what different models have in common differs in many respects from understanding what different metrics have in common.

In this light, it is significant that Wilber, like Dawson, ultimately suggests that what is common across different developmental approaches is best understood in terms of the methods and metrics behind the models. Wilber (2005, excerpt D) even goes so far as to say, with appeals to Kant, Peirce, and Habermas, that it should be possible to *rationally reconstruct* the "primordial perspective" that sets the conditions of the possibility for the methods used in developmental psychology. That is, if we turn away from the various models and toward the basic practices that make them possible we find a certain fundamental commonality. What I see in Wilber's meta-theoretical position—which I interpret in light of Dawson empirical efforts and Habermas's theory of the lifeworld—is the suggestion that there is some basic *intuition of altitude* woven into the fabric of the taken-for-granted practices and beliefs that constitute the lifeworld. This amounts to saying that, perhaps as a basic communicative or social capability, we cannot help but engage in making intuitive judgments about the relative development of individuals and artifacts. Moreover, this very basic capability frames and facilitates the methods of developmental psychological, which are in effect refinements and improvements on this perspective we always already enact in understanding others and ourselves.

On rational reconstructions

Habermas (1990, 1998) has suggested that the best way to understand the work of Piaget and Kohlberg is to construe their models as rational reconstructions of the intuitive knowledge mastered by subjects at different levels of development. I am taking Habermas's advice a bit differently than he intended. Instead of construing developmental *models* as rational reconstructions of knowledge as it develops in a particular domain (i.e. reconstructions of particular leaning sequences), I am suggesting that it is better to construe developmental *metrics* as rational reconstructions. That is, I am suggesting that we take one more step back and look at what allows us to disclose developmental differences between performances in the first place; i.e. what allows us to *see* development before we even begin to reconstruct the knowledge structures that constitute the various stages? In fact, I think that only after we have rationally reconstructed a metric can we move towards reconstructing learning sequences in particular domains. This way of approaching the issue is a bit more radical and results from contemporary efforts in cognitive development successfully triangulating domain general developmental assessment criteria (Dawson-Tunik, Commons, Wilson, & Fischer; 2005). I think this revision of Habermas's insight makes good sense.

The basic idea of *rational reconstruction* stems from Habermas's (1988) early preoccupations with the logic of the human sciences. He was out to justify the basic methods of psychology and sociology, fending offer behaviorist reductions, all the while keeping an eye fixed on hermeneutics and phenomenology. Thus assimilating the "interpretive turn" in social science methods he concluded that the most viable approach to theory and research involves "procedures that systematically reconstruct the intuitive knowledge of competent subjects" (Habermas, 1998 p. 34). He references, for example, Chomsky's efforts in linguistics, where the thing to be explained was our remarkably complex grammatical competence. Habermas suggests that Chomsky's theory of universal grammar is a rational reconstruction of the intuitive knowledge mastered by competent speakers. Even those who don't know the rules of grammar

can speak grammatically (and judge the relative grammaticality of statements made by others). He refers to other theorists, most notably citing Piaget and Mead, who employed "reconstructive procedures to transform a practically mastered pre-theoretical knowledge (*know-how*) of competent speakers into an objective and explicit knowledge (*know-that*)" (Ibid p. 35).

In the context of Chomsky's efforts this reference to Ryle's distinction between *know-how* (the unreflective ability to do something) and *know-that* (the explicit knowledge of how it is done) betrays the philosophical and normative edge of most rational reconstructions. That is, in the process of explicating how it is we are able to do something we often become more aware of the nature and structure of the task, allowing us to reflectively improve our practice. Despite our intuitive competence, learning the rules of grammar helps us speak and write grammatically and better judge the grammaticality of statements made by others. This means that rational reconstructions can serve a *critical* and *constructive* purpose (Habermas, 1990 p. 32). But this moment of idealism and abstraction in all rational reconstructions "should not discourage attempts to put rational reconstructions of presumably basic competencies to the test, subjecting them to indirect verification by using them as inputs into empirical theories" (Ibid).

With this brief account of Habermas's views and the example of Chomsky's efforts I am now in a position to state my basic hypothesis briefly before I go onto present empirical findings in its support. I propose that normally socialized humans possess a pre-theoretical knowledge allowing them to make accurate judgments about the relative development of the thought and language capabilities of others (akin, perhaps, to our intuitive ability to make judgments about the grammaticality of statements). This *intuition of altitude* (like our intuition of grammaticality) is an indelible aspect of the basic practices that constitute the lifeworld and can be explained in terms of certain invariant psycho-social structures implicit in interaction, communication, and socialization. The best way to understand developmental assessment systems is to take them as rational reconstructions of this intuitive competence (as the formal grammars constructed by linguists are rational reconstructions our intuitive sense of grammar).

Research

Background on the developmental sort-task

Inspired by efforts in the Kohlbergian tradition studying patterns of preference for arguments from different moral stages (e.g. Rest, 1971 and Walker, 1984; see also: Boom et. al, 2001) the developmental sort task was originally devised by Theo Dawson for purposes of training and explanation. In this capacity, the sort task has been used as an activity during workshops, lectures, and training sessions on the Lectcial™ Assessment System for almost a decade. This use of the sort task involves providing all participants in a group with a set of small cards, each with a short statement printed on it. Table 1 contains a sample of such statements, which exemplify the type of data that cognitive developmental psychologists typically collect. They are spontaneous linguistic performances gathered via clinical interview processes. In such interviews, subjects are prompted to focus their thought on some facet of some domain. The examples in Table 1 are from interviews conducted with individuals between the ages of 4 to 44, who have been asked to reflect on the domain of leadership. They are correctly sorted from youngest (1) to oldest (4).

Once card sets are distributed, group members are asked to *sort the examples developmentally* (i.e. in the order in which they think it would have been possible for a single individual to make the statements over the course of a lifespan). Then group leaders collect the sorts, display them to the group, and begin a discussion wherein members of the group explain why they sorted the statements as they did. To our initial surprise, there is *always* a remarkable degree of consensus regarding the sort. Moreover, the modal sort is *almost always* correct—as pre-determined by experts in developmental assessment. This finding has been replicated with hundreds of people in dozens of groups of various sizes and in extremely various contexts using a wide variety of examples (for example, asking subjects to sort as many as 15 examples from domains as diverse as physics and ethics).

Protocol #	Transcripts
4	[From your perspective, what are the qualities of a good leader?] I think a leader… is a person who reframes what is believed to be appropriate or possible. There's a lot woven into that statement. Leadership is service, which means helping people make meaningful sense of, and become personally involved with, a coherent sense or statement of what the organization's about and to guide them in the co-creation of a new statement or better statement, a refined statement that subsumes individual goals into collective goals. (ID: 0058)
3	[What you think makes a good leader?] Someone who follows the rules and sets boundaries. Someone who does what they have to do in order to move forward. They also try not to make a bad example out of what they are doing. Good attitude in all things. Nice personality, get to know everybody, helps everybody. Does not exclude anybody. A leader needs to let people know where they're going and know what they should or should not do. (ID: 10107)
2	[What makes a good leader?] If you can trust the leader. [And why is it important to trust the leader?] So you know that you don't have to lie or something. [Why is lying bad?] Because it can get you into more lies. [Ok…is there anything else you can think of about a good leader?] Being honest. [And why is that important?] Because you need to be honest. Because if you lie, something bad might happen. [What do you mean?] Like you can't stop lying. (ID: 0171)
1	[Okay, what do you think is a good leader?] It's someone who is in charge or in front of a line or something. [What would make someone a good leader?] Not screaming and crying. [Not screaming and crying, why is that?] Because then you would be making a fuss. You aren't supposed to make a fuss. It's bad. (ID: 1083)

Table 1: Four examples of reasoning in the leadership domain.

The discussions following the sort task activity are also intriguing. It has become clear that sorting the cards involves a series of *judgments* about the relative developmental levels of the various statements, and that individuals justify and explain these judgments in different ways. That is, individuals make reference to various properties of the statements that have taken to be indicative of development, including, for example, sentence length, logical soundness, egocentricity, or the experiences implied by the ideas expressed. As more and more groups displayed the ability to accurately complete the developmental sort task and reflect on the properties they took to be indicative of development—all with minimal prompts and without any training in developmental assessment—it became clear that these informal findings warranted further research.

Current research efforts

The results presented in this paper are a preliminary exploration of the trends in data collected from the first 181 subjects to participate in a web-based version of the developmental sort task called the Sort Task Challenge (STC). The on-line data collection interface involves a brief explanation of the research, a collection of basic demographic information, an explanation of the task (e.g. sort the given statements in the order in which you think it would have been possible for a single individual to make the statements over the course of a lifespan), and a set of ten statements in a pdf document, which participants can print out, cut apart, and sort by hand. The current STC is on the topic of leadership, and includes ten cards displaying statements of approximately equal length, and representing reasoning from six developmental levels (from Representational Mappings through Single Principles; see Figure 1 above). Because this study is still underway I will not display the statements used, and I have changed the statement ID letters on all tables. Participants are asked to match each card to one of ten ages (from 4 to 45). They are also prompted to write a brief explanation of the criteria they employed to sort the cards.

The 181 participants were volunteers, all of whom visited the Developmental Testing Service web site (www.devtestservice.com). Some were solicited, others earned from word-of-mouth or search engines. There was no organized formal effort to recruit participants. Table 2 displays the distribution of ages in the sample. We are collecting other demographic information, but in this initial and exploratory study, only the effects of age differences are explored.

Table 2: Distribution of ages in sample (N=181)

Under 20	54	**Youngest**	11
Between 20-30	71	**Oldest**	71
Between 30-50	28	**Mean (age)**	29
Over 50	25	**Mode (age)**	19

Analysis

Several questions and aims framed the initial exploratory analyses presented here. First, I want to characterize and describe the relation between the sorts offered by participants and the sort based on the actual developmental level of the statements (i.e. the relation between the *intuition-sorts* and the *metric-sort*). I did this by creating tables displaying the distribution of intuition-sorts in relation the metric-sort and developmental levels. Displays of basic descriptive counts are a useful way to show trends in the data (see tables 4-6). Also, following Boom et al. (2001), I calculated a Pearson Product Moment Correlation coefficient for each participant's intuitive-sort relative to the metric-sort, which indicates how well a particular intuitive-sort matches up with the metric-sort. These were used to calculate group average correlations, which represent how well, in general, the intuitive-sorts of a particular group members match up with the metric-sort. I computed group average correlations for the whole group (N=181), for participants under 20 years of age (N=54), and for those over 50 years of age (N=25).

I was also interested in the participants' explanations of their intuitive-sorts. To begin, I read through each participant's brief explanation, and generated a set of four grounded coding categories. Explanations and justifications of intuitive-sorts were categorized as, *descriptive*, *objective*, *evaluative*, *instinctive*, or some combination of these. ***Descriptive*** explanations were those in which the participant referred to their working model of the domain or life-course; i.e. they appeared to have a mental model of development and judged the relative developmental level of statements in light of it. For example, a participant whose explanation was coded as *descriptive* might suggest that young kids only know certain kinds of leaders, while adults come to have more experience with leaders and organizations; or, that development proceeds from egocentric to ethnocentric to world centric. These explanations are characterized by their relative domain specificity and by their allegiance to some substantive hypothesis or rich description of the life course. ***Objective*** explanations were those in which participants referred to some universal property of all concepts or linguistic performances. A participant whose

explanation was coded as *objective* might have referred to such things as vocabulary, sentence length, complexity, or integration, etc. These types of explanations are characterized by their domain generality and minimal theoretical/descriptive commitments. ***Evaluative*** explanations were those in which the participant referred to the greater value, validity, or adequacy of higher-level responses—they sorted in order of preference. These types of explanations are characterized by the way the value judgments of participants determine the sort. Finally, ***instinctive*** explanations were those in which participants simply said they "used their gut" or "it seemed right to me."

These categories were then used to code the explanations provided by participants.

Results

Table 3 presents the relation between the intuition-sorts and metric-sort for the full sample. There are several things to note. First, following the bolded boxes diagonally from the bottom left to the upper right, we can see how often the mode of intuition-sorts for the group (in red) coincides with the metric-sort. The trend is clear. For all but the top two levels, the *majority* of participants put the statements in the right position. It is also worth noting how the distribution of intuition-sorts becomes increasingly "noisy" toward the upper levels. Both of these trends point toward tentative conclusions that are addressed in the discussion section.

Table 3: Displays distribution of intuition-sorts on developmental sort task for whole group (N=181). Red text represents mode sort for group. Bolded boxes represent the correct sort.

Rank order	Card ID in order (level)	q	r	s	t	u	v	w	x	y	z
10	**z** (SP)				1	2	4	16	13	89	55
9	**y** (SP)			1	2	2	8	32	28	39	56
8	**x** (AS)				4	5	4	57	65	21	23
7	**w** (AS)				2	9	15	61	47	15	30
6	**v** (AM)				5	52	76	14	12	9	10
5	**u** (AM)			2	9	82	54	1	4	5	3
4	**t** (SA)	1	2	7	127	25	17				
3	**s** (RS)		52	118	8	2	1				
2	**r** (RS)	2	122	54	2			1			
1	**q** (RM)	178	3								

Tables 4 and 5 should be read in the same way. These tables display the trends in two age groups (under 20 and over 50). Comparing them allows us to make certain interesting observations. Most obvious is the fact that the older group is more accurate. The *majority* of older participants put statements in the right order more often than the majority of younger participants. Also, as we will see when we look at the average correlation coefficients for the two groups, there is more noise in the younger sample. The *majority* of participants under 20 placed statements inappropriately four times. Moreover, these errors begin at a lower level than comparable errors made by the older group and the full sample. These comparisons point toward intriguing conclusion about our intuitions of development, which are discussed below.

Table 4 and 5: Top: Distribution of intuition-sorts on developmental sort task for those over 50 years old (N=25). Bottom: Distribution for those under 20 years old. Red text represents mode sort for group. Bolded boxes represent the correct sort.

Rank order	Card ID in order (level)	q	r	s	t	u	v	w	x	y	z
10	**z** (SP)							1	1	10	14
9	**y** (SP)				1		3	3	5	5	9
8	**x** (AS)					1		9	10	3	2
7	**w** (AS)				2	2	2	9	6	4	1
6	**v** (AM)				1	6	10	4	1	4	
5	**u** (AM)				1	13	10		2		
4	**t** (SA)	1		1	20	3	1				
3	**s** (RS)		6	18	1	1					
2	**r** (RS)		18	7			1				
1	**q** (RM)	25	1								

Rank order	Card ID in order (level)	q	r	s	t	u	v	w	x	y	z
10	**z** (SP)						2	5	3	34	9
9	**y** (SP)					1	2	15	10	4	21
8	**x** (AS)				1	2	1	17	16	6	10
7	**w** (AS)					4	5	13	17	5	9
6	**v** (AM)				5	15	19	4	5	2	3
5	**u** (AM)			1	11	24	15		1	1	
4	**t** (SA)			2	33	8	9				
3	**s** (RS)		18	33	3						
2	**r** (RS)		36	18							
1	**q** (RM)	54									

Table 6 displays the average Pearson Product Moment Correlation coefficients for each of the three groups. Correlations are high for all groups, but older participants were more accurate than the full sample, which was more accurate than participants under 20. Importantly, I have not disattenuated these correlations for error, nor have I calculated their statistical significance. These numbers are merely descriptive and suggestive.

Table 6: Average Pearson Product Moment Correlation coefficients between intuitive rank-orders and actual orders for all sample groups

Sample Group	PPMC
Whole group (n=181)	0.86
Under 20 (N=55)	0.83
Over 50 (n=25)	0.90

The results of the analysis of participant's explanations of their intuitive-sorts are displayed in Table 7. Across the board, *descriptive* explanations are the most common. In the full sample, *objective* explanations are the second most common, followed by *evaluative*, and then *instinctual*. This same trend characterizes the explanations of participants over 50, although it should be noted that there is only a small difference (literally one participant) between the *descriptive* and *objective* categories. Trends for participants under 20 differ from those of the full sample and the older group. In this group, *evaluative* explanations follow hard upon *descriptive* ones, and *objective* and *instinctual* explanations are far less populated, relatively speaking. When we hold these results in mind, along side trends concerning differences in accuracy between the older and younger groups, an interesting picture emerges.

Table 7: Frequency and percentage of justification types for rank-orders for whole group, those of 50, and those under 20. Justifications that fell in more than on category were counted in each category in evidence.

Justification type	Whole group (n=181)	Under 20 (N=55)	Over 50 (n=25)
Descriptive: sort in terms of assumed working model of development or domain	110 (60%)	25 (45%)	13 (52%)
Objective: sort in terms of universal properties of language and concepts	40 (22%)	6 (11%)	12 (48%)
Evaluative: sort in terms of increasing preference for or value of responses at higher levels	27 (15%)	20 (36%)	2 (8%)
Instinctive: sort in terms of unexplained insight	15 (8%)	4 (7%)	0

Discussion: developmentally assessing others is a capability (that develops)

Despite their preliminary character, the results presented above sketch an intriguing picture and lend support to my basic hypothesis. I think the sort task gives credence to the idea that individuals are in possession of a robust pre-theoretical knowledge concerning the development of thought and language, which in this instance allows them to accurately sort statements developmentally. Assuming that the results obtained above hold beyond the leadership domain—which we have good reason to expect in light of the documented accuracy of sorts administered in a wide variety of domains—it appears that this is an *intuition of altitude.*

It is important to note both that this intuition becomes less accurate as it is applied to increasingly developed performances and that it appears to be a capability that develops, i.e. the older group was more accurate than both the whole group and the younger group. These two trends are suggestive, although it is hard to draw definitive conclusions at this time. Comparable studies in the Kohlbergian tradition (Rest, 1971; Walker, 1984; Boom et. al, 2001) have consistently shown a pattern where participants reliably judge the sophistication of arguments up to their own stage but become increasing unreliable at judging arguments beyond their own stage. Importantly, these studies were conducted so that all participants engaged with a set of statements in *evaluative* terms, i.e. participants were asked to perform preference-based rankings. It is worth noting that this approach not only betrays a basic theoretical

assumption about normative issues in developmental psychology (i.e. that a developmental sort *is* a preference-based sort); but prompting participants to perform preference-based sorts also appears as a methodological liability, as I have just shown that sorts justified in *evaluative* terms were more prevalent in the group that was least accurate. Nevertheless, it seems safe to say that capabilities for assessing development change over time, becoming increasingly accurate across a wider developmental range.

Participants' explanations of their sorts are also interesting. As just noted, sorts justified in *evaluative* terms were more prevalent in the group that was least accurate. It also appears that sorts justified in *objective* terms were more prevalent in the group that was most accurate. Overall, however, most participants made their judgments in light of a working model or rich description of the life-course. That is, most participants appealed to a rough sense of how development unfolds, and judged statements relative to this intuitive developmental model. Clearly, more work needs to be done to explore how different types of explanations are correlated with differing degrees of accuracy.

These findings about the way participants *unpack* their intuitions are also of *theoretical* interest because they are the everyday equivalent of rational reconstructions. These types of ad-hoc explanations of intuitive know-how are exactly the place where the *critical* and *constructive* power of rigorous rational reconstructions can gain leverage to effect the rationalization of lifeworld practices. Empirical and theoretical reconstructions of the know-how behind basic capabilities can be fed back into the lifeworld to function as reflective and normatively charged idealization of everyday accomplishments. This is how Habermas sees his own work on the formal pragmatics of communication and the theory of democracy and how he sees reconstructions of *the moral point of view* as offered by Kohlberg and Rawls. These reflections lead us to the implications of conceiving developmental assessments as rational reconstructions of our intuitions of altitude.

Conclusion: Intuitions of altitude and the future of developmental assessment

Under the banner of a post-metaphysical philosophical orientation, Peirce directed his attention to the conditions that allow us to get on with inquiry. Therefore, metrics were on his mind. His measurements of the Milky Way were based solely on a metric that was devised by empirically explicating certain previously taken-for-granted properties of his own eyes. At the outset I drew an analogy between Peirce's attempts to clarify the reliable differential responsiveness of his eyes and my attempts here to clarify our intuitions of altitude; they are both attempts to empirically establish the conditions of the possibly for certain measurements and methods. The analogy breaks down in so far as we must categorically differentiate measurements based on sensory experience from the measurements based on communicative experience (Habermas, 1988, p93). I clarified that the latter required procedures of *rational reconstruction*. But the analogy is useful nevertheless, in part because it exposes the fact that the roots of scientific methods run deep into the structures of the lifeworld and, ultimately, of our organism.

I have presented empirical evidence to support my contention that a large sub-set of developmental assessment systems can be traced back to a common intuition, which I propose forms an indelible part of the practices that constitute our shared lifeworld. For example, most forms of communication, but particularly the language games of teaching, parenting, and counseling, require an ability on the part of participants to gauge the relative development of the thought and language of interlocutors in order to communicate effectively. As I suggested above, just as we make intuitive assessments of the grammaticality of statements made by ourselves and others, so we make intuitive assessments about development. Everyday observations bear me out in so far as all cultures have a fairly nuanced set of adjectives for charactering the intelligence, sophistication, or maturity, of individuals. I cannot think of a time or place in history when the use of intuitive assessments of these qualities would not have formed an important part of formal and informal social, cultural, and institutional norms.

I also cannot think of a time or place in history when these informal assessments were carried out with explicit concern for objectivity and fairness. That is, although we always already engage in making developmental assessments of others, we do not all share a common metric. We do not apply the same set of intersubjectively codified criteria (as the sort-explanation typology revealed above) and we cannot all assess the same set of performances with equal accuracy (as the "noise" in the data demonstrates). This is where the *critical* and *constructive* function of rational reconstructions becomes important. I am suggesting that formally constructed and empirically validated developmental assessment systems improve upon on the informal assessments that shape our everyday lives. As such they *ought* to replace informal assessments in certain contexts, like schools, businesses, and government. If we have a ruler on hand why should we rely on intuitive measurements in important situations? And while I do not claim that everything that is relevant in making assessments of persons is amenable to rational reconstruction, I do think that we have succeeded in codifying a metric based on at least one ubiquitous index of development (Dawson, 2008; Stein & Hiekkinen, 2008).

The notion that developmental assessments should be understood as rational reconstructions of an intuition of altitude also opens up the field to a kind of *methodological pluralism* (perhaps in this case it is a *metrological pluralism*). There are many ways to unpack an intuition. Just consider the various formal grammars following in the wake of Chomsky, all aimed at explicating one and the same intuition of grammaticality. Or consider the various valid metrics for heat and length; different systems of measurement but all fundamentally based on invariant relations between the organism and environment. Of course, different metrics are good for different things. I'm suggesting that choosing the right metric requires we adopt a *problem-focused metrological pluralism* (Dawson, Fischer, & Stein, 2007) where we look across the various metrics available and employ the one most fitting, sometimes trying a few out before settling on the best one. And if we are honest about the breath of human experience and capability, it's safe to say we will need to build new metrics as new interests emerge.

Sources:

Baldwin, J. M. ([1906] 1975). *Thought and things: A study in the development of meaning and thought or genetic logic* (Vol. 1-4). New York: Macmillan Co.

Boom, J., Brugman, D., & van der Heijden, P.G.M. (2001). Hierarchal Structure of Moral Stages Assessed by a Sorting Task. *Child development*, 72 (3), 535-548.

Dawson, T. L. (2008, 6/22/04). The lectical™ assessment system. 1. Retrieved June, 2008, from http://www.lectica.info

Dawson, T. L. (2001). Layers of structure: A comparison of two approaches to developmental assessment. *Genetic Epistemologist, 29*, 1-10.

Dawson, T.L. (2002). A comparison of three developmental stage scoring systems. *Journal of Applied Measurement, 3(2),* 146-189.

Dawson, T.L. (2004). Assessing intellectual development: Three approaches, one sequence. *Journal of Adult Development, 11(2),* 71-85.

Dawson, T. L., Fischer, K. W., & Stein, Z. (2006). Reconsidering qualitative and quantitative research approaches: A cognitive developmental perspective. *New Ideas in Psychology, 24*, 229-239.

Dawson-Tunik, T. L., Commons, M. L., Wilson, M., & Fischer, K. W. (2005). The shape of development. *The International Journal of Cognitive Development, 2,* 163-196.

Fisch, M (1986). Introduction. In Peirce edition project (Ed.), *Writings of Charles S. Peirce: A chronological edition* (Vol. 3). Bloomington, IN: Indiana University Press.

Fischer, K., & Bidell, T. (2006). Dynamic development of psychological structures in action and thought. In W. Damon & L. R.M. (Eds.), *Handbook of child psychology: Theoretical models of human development* (Vol. One, pp. 1-62). New York: John Wiley & Sons.

Fisher, W.P. (2004) Meaning and method in the social sciences. Human Studies. 27, 429-454.

Fisher, W.P. (2005) Mathematics, measurement, metaphor, and metaphysics: implications for method in post-modern science. Theory & psychology, 13(6), 753-790.

Habermas, J. (2001). *On the pragmatics of social interaction* (Fultner, Trans.). Cambridge: MIT Press.

Habermas, J. (1998). What is universal pragmatics? (M. Cooke, Trans.). In M. Cooke (Ed.), *On the pragmatics of communication* (pp. 21-105). Cambridge, MA: MIT Press.

Habermas, J. (1992). *Post-metaphysical thinking: Philosophical essays*. Cambridge, MA: MIT Press.

Habermas, J. (1988). *On the logic of the social sciences* (Nicholsen & Stark, Trans.). Cambridge: MIT Press.

Habermas, J. (1990). Reconstruction and interpretation in the social sciences (Nicholsen, Trans.). In *Moral consciousness and communicative action* (pp. 21-43). Cambridge, MA: MIT Press.

Husserl, E. (1970). The crisis of european sciences and the transcendental phenomenology. (Carr, Trans). Northwestern University Press. Evenston.

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 347-480). Chicago: Rand McNally.

Peirce, C.S. (1868). Questions concerning certain faculties claimed for man. In Peirce edition project (Ed.), *Writings of Charles S. Peirce: A chronological edition* (Vol. 2, pp. 193-211). Bloomington, IN: Indiana University Press.

Peirce, C. S., & Jastrow, J. (1885). On small differences of sensation. In Peirce edition project (Ed.), *Writings of Charles S. Peirce: A chronological edition* (Vol. 5, pp. 122-135). Bloomington, IN: Indiana University Press.

Peirce, C. S. (1878). Photometric researches made in the years 1872-1875. In Peirce edition project (Ed.), *Writings of Charles S. Peirce: A chronological edition* (Vol. 3, pp. 382-493). Bloomington, IN: Indiana University Press.

Peirce, C.S. (1882). Introductory lecture on the study of logic. In Peirce edition project (Ed.), *Writings of Charles S. Peirce: A chronological edition* (Vol. 4, pp. 378-382). Bloomington, IN: Indiana University Press.

Piaget, J. (1972). *The principles of genetic epistemology* (W. Mays, Trans.). London: Routledge & Kegan Paul.

Kuhn, T.S (1977). The function of measurement in modern physical science. In *The essential tension: selected studies in scientific traditions and change* (pp. 178-224). Chicago: University of Chicago press.

Rest, J. R. (1973) The hierarchal nature of moral judgment: a study of patterns of comprehension and preference of moral stages. *Journal of Personality*, 41, 86-109.

Stein, Z & Hiekkinen, K. (2008). On operationalizing aspects of altitude: an introduction to the Lectical Assessment System for Integral researchers. *Journal of integral theory and practice*. Spring. Vol 3. No 1.

Stein, Z. (in preparation) Dissertation research. Harvard University Graduate School of Education. Cambridge MA.

Walker, L.J., de Vries, B., & Bichard, S.L. (1984) The hierarchical nature of stages of moral development. *Developmental psychology*, 20, 960-966.

Wilber, K. (1995). *Sex, ecology, spirituality - the spirit of evolution*. Boston: Shambhala Publications.

Wilber, K. (1999) Integral Psychology. Boston: Shambhala Publications

Wilber, K. (2005). Excerpts from Volume 2 of the Kosmos Trilogy, forthcoming: Integral Institute.