

It's an Empirical Question: On Cognition and Ego

Katie Heikkinen

In light of the wonderful rejoinders to the paper by Zachary Stein and myself, I offer a few informal (and rather rushed) comments in response, focusing on three elements: 1) the dangers of labeling colleagues; 2) psychometrics; and 3) the relationship between cognition and ego. In the second half of my paper, I describe my current research project, which I believe will be of interest to many of you.

On Labeling Colleagues

One point made by Stein in his paper (this volume) is of such great importance that I wish to restate it here: “disqualifying” certain arguments as coming from a lower developmental position is both illogical—because this stance uses the “truth” of a theory to negate critiques of that supposed truth—and unethical—because this stance devalues and marginalizes voices of opposition. But I’d like to extend his point to note that while “developmental disqualification” is a rather egregious sin, using developmental theory to label colleagues, even if in good fun, is nearly as egregious.

For example, Stewart, Smith, and Suzuki (this volume) quip that “we can all acknowledge a strong ‘T/Achiever/Orange’ streak” in some but not all of our colleagues. Since everyone he mentioned are making psychometrics for the marketplace and, I presume, care deeply about human development, I struggle to see the grounds for distinction here. We are all in the same boat! It seems to me that the distinction hinges on some rather messy confluences of structure and content. Various values such as interest in science or desire to self-promote (the content) can be held at multiple levels of complexity (the structure). This is an important distinction that Stein and I have made elsewhere (2008).

Without clear referencing to adequate data—and I think referencing particular value content is neither adequate nor clear—these labels are meaningless. They only serve to antagonize and threaten. Is this the kind of community we want?

On Psychometrics

Stein and I made a bold claim that the LAS and HCSS are the only calibrated metrics (2009). I am extremely grateful to Torbert and Livne-Taranadach (this volume) for their detailed description of the reliability testing of the LDP; to Stewart, Smith, and Suzuki (this volume) for their call for greater inclusion of the practitioner’s perspective; and to Basseches (this volume) for highlighting the importance of philosophical argument in determining the desirability of certain developmental outcomes. However none of these speak directly to the issue of calibration. While evidence of blind interrater reliability studies goes part of the way, it is not sufficient evidence of reliability to satisfy the requirements for individual assessment, which demand that statistical reliability is high enough to allow us to have a certain level of confidence



in each possible score (See Dawson's table, this volume). Both Torbert and Cook Greuter argue that they have scored hundreds, if not thousands of SCT's. Submitting their data to a Rasch analysis to examine the statistical reliability of their scales would be a simple matter.

Finally, I wish to point out a few questions that remain open after our publication and its rejoinders. Has anyone further evidence of validity or reliability studies of the Spiral Dynamics metrics? What are the implications of Jane Loevinger having said the test was not to be used to measure individuals (Loevinger, 1979; 1993)? What is the "right relationship" between first, second, and third person approaches (Torbert and Associates, 2004) in the field of psychometrics?

On Cognition and Ego

In this issue, Murray wonders whether "the structural transformations happening inside the body (or body/mind/soul) happen according to a different type of mathematics or logic or geometry" than the logic of hierarchical complexity (p. 344). This is a very reasonable musing that I myself, and many of my students at John F. Kennedy University, have made. I agree with his observation that change in such domains as compassion or inner peace *feel* very different than change in understanding, say, physical science concepts.

However, thus far we have no evidence of this alternative logic. To unfold may in fact *mean* to unfold through hierarchical complexity. Note that I do *not* mean that hierarchical complexity is the only important or real dimension of human change. To the contrary—there are many, many human change trajectories—some healthy and adaptive, others less so. But to conflate change—even the healthiest, most adaptive changes of them all—with development is problematic (e.g. Mascolo and Fischer, in press). As Stein and I have pointed out elsewhere (2008), linking a notion of development with particular content—such as Murray's examples of attributes that characterize wisdom, from "letting go" to a "sensation of expansion" to comfort with "dissonance, uncertainty, ambiguity, and paradox" (p. 344)—makes it impossible to see high forms of development that are not characterized by that content. Surely there must be someone who is highly developed—a complex thinker skillfully adapted to his world—who does not feel expansive?

Ah-ha, I can hear you thinking. *But, Katie*, that means *that person is not developed*. Yes, but only if you define "development" as "expansiveness." But as Stein (this volume) notes, to do so is to make a normative move. Implicit in that distinction is something like "I believe that it is better to be expansive than non-expansive." This may very well be true! But then we must 1) be honest that we are making evaluative claims; 2) be able to explain clearly why we view such changes as developmental—without making recourse to explanations inherent to the claim being made (e.g. "it is higher because it is higher"). For example, one might argue that expansiveness is developmental simply because the progression is hierarchical. That is, the end state of "expansiveness" hierarchically integrates lower order tasks (even if by exclusion): expansive pre-contraction (e.g. $A = \{e\}$), expansive with contraction (e.g. $A = \{e, c\}$), and expansive post-contraction (e.g. $A = \{e, \text{not } c\}$).

This is fine—a reasonable definition of development that does not define itself by itself (instead it calls on structural distinctions). But how could we *measure* such a concept? I'll exclude for now methods like clinical observation, simply because the Stein and Heikkinen paper focused on more formal metrics—so we are left with something like self-report. For example, we could ask people to explain their emotional and embodied states in an open-ended interview. If we then scored these interviews using the LAS, we could find one of several trends: 1) The feeling of “expansiveness” is one that is *only* reported at higher Lectical levels; 2) The feeling of “expansiveness” is one that is reported at many developmental levels; 3) That some people reasoning at high Lectical levels don't mention expansiveness at all.

Each of these findings has different implications. #1 suggests that expansiveness as a feeling-state is only experienced or spoken about when people have developed a complex capacity to discuss their states overall. Thus it could be said that “expansiveness” is one of the characteristics common to highly complex performance of emotional state understanding. #2 suggests the opposite; that expansiveness is a feeling state available to people operating at many levels of complexity. So the concept doesn't characterize a particular level; instead, we can track whether and how the concept changes with development. What does a simple understanding of “expansiveness” look like, compared to a more complex understanding? #3 is an almost certain finding, whether #1 or #2 prove true. It is rare for *every* sample at a given complexity level to contain the same concept. So some people who reason about their feeling states in a highly complex—highly developed—way do not talk about a sense of expansiveness. We might argue that this person is less healthy or behaves less adaptively, but in this view, we cannot say they are less developed.

Another option might be to score these self-reports with a newly developed scoring system, based on the surface structure of the performances. However, if this new system were anything like Kohlberg's, Armon's, or Perry's, the LAS would serve just as well (e.g., comparison with Kohlberg: Dawson, 2003; comparison with Armon: Dawson, 2002, Dawson-Tunik, 2004; comparison with Perry: Dawson, 2004).

Some of you may still wish to argue that expansiveness is not necessarily coupled in this way to the capacity to reason about it linguistically—that regardless of self-report, “I know it when I see it” or “I know it when I feel it.” This is reasonable. (First of all, I must make the caveat that “hierarchically complex” at least in terms of the LAS does not mean “complicated.” It means highly differentiated and *integrated*. So the LAS does not miss the “simplicity on the other side of complexity,” as some might claim.) But, again, how can we *measure* that ineffable state of being? We might use clinical assessments based on a set of criteria. A fine approach, but since it weighs on the expertise of the clinician, this approach is less scalable. We might instead develop a trait-inventory similar to a personality test. Fine again, but that is not developmental. At the simplest level, we can watch people and feel into them to determine how expansive they are. However, without psychometric procedures, this allows us to say nothing about the reliability or validity of such a judgment, or whether the capacity develops.

Note that in the end, even if we believe that the capacity to be or feel expansive is not tied to the capacity to reason about it linguistically, that it is somehow different and unfolds differently, we are still left with scant options other than linguistic self-report—if we wish to make strong

claims about it being a reliable, valid, and developmental psychological construct. But if we do not wish to make such claims, clinical assessments, trait inventories, or informal “grokking” all have their place. In this light we might think of “expansiveness” as something akin to “kindness:” a valued character trait where most people “know it when they see it” but that has no psychometric or developmental properties. This doesn’t lessen the importance of kindness—norms matter deeply, but they should not be part of developmental metrics.

What the LAS has attempted to do is strip the normativity away from the assessment process. Stein (this volume) notes that such notables as Baldwin, Piaget, Kohlberg and Sullivan “used prescriptive, normative, or ethical language, [but] they did not draw it directly from the empirical substance of their models” (p. 15). This point is worth repeating—*they turned to philosophers*. I don’t think that this implies that psychologists have nothing to say about the normative aspects of human development. Some psychological findings—such as those linking various behaviors to happiness or mental health—do offer insight into how humans might attain certain desirable outcomes. But determining whether those desired outcomes *ought to be desired* is not the job of the psychologist—unless they are clear that they are “putting on their philosopher hat” in so doing.

I hope I’ve highlighted some of the challenges in determining whether “the structural transformations happening inside the body (or body/mind/soul) happen according to a different type of mathematics or logic or geometry” than the logic of hierarchical complexity (Murray, this volume, p. 344). Perhaps it’s true! But the science of measurement, it seems, struggles to make this distinction. Whether this is a current struggle—reconcilable with changes in methods—or an intractable struggle—because these aspects can only be known through linguistic performance—remains to be seen.

My Current Research Project

In this issue, Murray wonders whether “ego development” metrics such as those from the Loevinger tradition are “assessing some of the wisdom skills that [he] argue[s] may have elements that do not develop according to hierarchical logics” (p. 345). Musings about the relationship between ego development and hierarchical complexity are a start, but this is fundamentally an empirical question. One way to address this question is to re-score ego development data using a structural method (the LAS or the MHC). However, this is hard to do with SCTs because the stems are so short. Short sentences can be scored by the LAS, but without any justification by the participant, less confidence is placed in the scores. However, another such “ego development” model is Robert Kegan’s constructive developmental theory (1982; 1994), which features a validity profile similar to the current versions of the SCT (e.g. Torbert et al, this volume) in that it focuses on external or predictive validity (e.g. Kegan, 1994; Harris and Kuhnert, 2008; Lewis et al., 2005). I am currently working on a research project to address just this empirical question. I’d like to share my thinking with you here, in hopes that it sheds more light on some of the important distinctions Stein and I have made in previous publications (2008; 2009): content versus structure and metric versus model.

Kegan’s constructive-developmental model details how a person’s meaning-making structure undergoes qualitative shifts over the lifespan. This model includes five stages of change in an

individual's "order of mind," which is an overarching structure of self-system organization. An "order of mind" includes cognition as well as intrapersonal and interpersonal abilities, all of which are hypothesized to change in concert with one another. Individuals are located on the stage spectrum by participating in a "subject object interview" (or SOI) where they discuss events that are emotionally significant to them. These interviews are transcribed and then scored using an inductive process that is supported with examples. The scoring process also depends on the skill of the interviewers, since they develop and test stage hypotheses during the interview itself (Lahey, Souvaine, Kegan, Goodman, and Felix, 1988).

Relying on an inductive process—where scorers infer underlying structure of meaning making—has important advantages over scoring methods that rely more heavily on concept-matching strategies, such as Kohlberg's (Colby and Kohlberg, 1987). Concept matching strategies "suffer from bias introduced by nonrepresentative construction samples and an overdependence on particular content" (Dawson, 2003, p. 337). Inferring underlying structure allows raters to "look through" disparate-seeming surface content to see similar structures that lie beneath (or vice versa). However, I argue here that the SOI, like the Standard Issue Scoring System, can only claim to "peer beneath" one layer of *structure*—the conceptual *content*—to view the surface (or domain) structure of a performance (Stein and Heikkinen, 2008). This places it in fine company—this is true of nearly all extant developmental metrics.

In contrast, the Hierarchical Complexity Scoring System and the Lectical Assessment System claim to be able to peer beneath *both* conceptual content and domain structure to discern the core structure that lies beneath. In recent years, research by Dawson and colleagues has shown that interview data collected for analysis with various domain-specific scoring systems can be scored with the domain-general Lectical Assessment System (or its precursor the HCSS), yielding results that suggest that all of these systems measure the same underlying developmental dimension, hierarchical complexity (e.g., comparison with Kohlberg: Dawson, 2003; comparison with Armon: Dawson, 2002, Dawson-Tunik, 2004; comparison with Perry: Dawson, 2004; see also Kitchener and Fischer, 1990 on reflective judgment). Dawson's research suggests that this single dimension is indicative of development in *any* domain about which we can reason.

This claim has important implications for the measurement of development, including for systems like the SOI that claim to assess "whole people." Might metrics of ego development less directly measure the same latent dimension? A handy, though limited, analogy is the importance of controlling for socioeconomic status (SES) in multiple regression studies. Since we know SES impacts so many educational outcomes, we could not trust our results if we did not statistically control for SES. Similarly, since hierarchical complexity is likely a latent dimension of *all* verbal performances, we must assess it separately in order to tease apart its contribution to the performance from the contribution of other factors.

By separately assessing hierarchical complexity, we are able to ask other important questions of the data. Just as controlling for SES allows the impact of other variables of interest to be seen, "controlling for" hierarchical complexity allows us to independently analyze conceptual change in the domain. We can see more clearly which aspects of the performance are related to an increase in hierarchical complexity and which aspects are not. The latter are referred to as "horizontal" differences (Dawson, 1998)—attitudes, dispositions, preferences, factual

knowledge, and biases that vary at least somewhat independently from “vertical” complexity. Parsing out these two dimensions allows us to take a closer look at individual difference. We are also able to examine which concepts arise at what Lactical level (because concepts per-se are not tied to a specific level) and trace how concepts that arise early and persist change with development. Tracing a concept through Lactical levels produces a developmental sequence or set of learning pathways that is open to continual refinement (Dawson-Tunik, 2004; Dawson and Stein, 2008).

By scoring the SOI transcripts using the LAS, I will be able to discern how much of the variance in SOI scores is explained by hierarchical complexity. Note, however, that the claim is *not* that hierarchical complexity can explain all the important dimensions of human development. On the contrary, Dawson stresses that hierarchical complexity is *only one* dimension by which to evaluate performances. Her methodology of “developmental maieutics” calls for the close partnership of scoring experts (Lactical analysts) with content experts (experts in the domain being assessed) to ensure that that the full richness of the domain is included (Dawson and Stein, 2008). However, the analysis of domain specific content is a separate process, distinct from the process of scoring. Only after both scoring and content analyses have been completed are the scores and themes re-married to produce rich descriptions of conceptual development—the developmental sequences.

My study will reveal whether interviews like the SOI where people talk about their emotional reactions to important events in their lives provide good “fodder” for the LAS. If the LAS scores and the SOI scores are in agreement, or if the LAS explains a large portion of the variance in scores, it suggests that the SOI coding system, like those of Kohlberg, Armon, and Kitchener and King, taps the hierarchical complexity dimension. This may be due to a property of the scoring method itself—the SOI manual may actually take hierarchical complexity into account in its process, despite calling it by another name. Or, it may be a completely implicit effect—more hierarchically complex samples simple “feel” higher, regardless of the explicit scoring instructions.

This finding would raise questions about the need for a distinct scoring system for ego development (Theo Dawson, personal communication, 11/14/09). The SOI method of scoring would be redundant if it did not add any additional information over and above hierarchical complexity. This does not mean that self-object theory itself would necessarily be brought into question. In fact, results of this kind would serve as an external validation of the current sequence. But using the LAS to score SOI performances would allow researchers to continue to address, in a rigorous manner, the empirical question of how the content of performances on the SOI change with development. As I’ve mentioned earlier, separately analyzing hierarchical complexity and conceptual content allows us to create learning sequences that carefully track the evolution of a concept or skill. These learning sequences themselves are free to change as the research base accumulates or as the culture changes and how concepts are used is altered. In other words, the scale stays the same even as what we know about development in that domain changes. In addition, clarity around what elements of performances are vertical and which are horizontal allow for the horizontal dimensions to be studied with other methods, like the thematic coding in the present study—or myriad possibilities like surveys or scales. Finally,

using the same scoring method and scale in multiple domains allows us to make meaningful comparisons across domains, which is not possible when each system's score stands alone.

An important distinction to remind the reader of here is between *model* and *metric* (Stein and Heikkinen, 2009). Both models and metrics in developmental psychology are *representational devices*: “methods, symbol systems, and propositions—built by communities of developmental psychologists—which in some way claim to be *about* cognitive developmental processes” (p. 7). Metrics attempt to assess how much of a psychological trait is present, while models attempt to explain its presence and describe its origin. The LAS and the SOI are metrics; skill theory and constructive developmental theory are models. While models are necessarily linked to their respective metrics, they are not the same. In the present study, the implications impinge on the SOI metric, not the CD model. The subject-object shift may be an important causal mechanism in how development occurs—but hierarchical complexity measures this shift just as well, or as Dawson has argued, even better.

My study will also reveal the role of hierarchical complexity in how people talk about important events in their lives. Complex thinking may play little role in the interview as a whole—complex language is simply not needed to tell a good story about an important event. But the little moments where people show evidence of their more complex ways of thinking may say a lot. These “moments of complexity” may provide insight into how the performance as a whole hangs together; and indeed shed light on their “meaning making structure” by explicating how they perceive interconnections and links between various elements of their story. The learning sequences in particular might reveal how the concept of (for example) “taking responsibility for my own emotions” changes across development—from a “hollow concept” that is stated but not explained to a more rich and nuanced description of what such responsibility means and what forms it should take. These sequences may, in the end, echo Kegan's writings, but they will be arrived at via a different method. This may provide additional support for certain aspects of the constructive-developmental model.

If the LAS cannot be used to score these data, there are several possible explanations. It could be a simple “mechanical failure”—the present data may lack certain properties that are required for scoring with the LAS (such as sufficient evidence of reasoning and adequate length of reasoning performances). But if there is enough material present to assign a firm complexity score, and the scores provide a different ordering of the passages than did the original rater, the implications are quite intriguing. The SOI may be looking for a property that does not vary with hierarchical complexity, but is on its own, independent developmental trajectory. Or, particularly at the higher levels of ego development, a new simplicity may be attained that hierarchical complexity fails to see. These are all questions Murray poses, but this kind of empirical approach has the potential to answer them.

Overall, my intention is to improve measurement in developmental science and to better bridge the gulf between ego and cognitive development theories. I think that understanding stage development in adulthood is important and useful and I encourage wide-ranging use of developmental thinking. But I also wish to see the *science* of developmental stages advance, and this involves building bridges between different measures and metrics. The LAS has the unique potential to be a domain-general *ruler* of any development that is expressed through language

broadly construed. Although other systems may tap into important aspects of development in their own right, only by accounting for hierarchical complexity can we make meaningful comparisons across systems and thus advance our understanding of development in multiple domains.

I am excited about this project and welcome any feedback or further inquiries.

References

- Colby, A., & Kohlberg, L. (1987). *The measurement of moral judgment, vol. 2: Standard issue scoring manual*. New York: Cambridge University Press.
- Colby, A., & Kohlberg, L. (1987). *The measurement of moral judgment, vol. 1: Theoretical foundations and research validation*. New York: Cambridge University Press.
- Dawson, T. L. (1998). "A good education is ... "A life-span investigation of developmental and conceptual features of evaluative reasoning about education. Unpublished doctoral dissertation, University of California at Berkeley, Berkeley, CA.
- Dawson, T.L. (2002). A comparison of three developmental stage scoring systems. *Journal of Applied Measurement, 3*(2), 146-189.
- Dawson, T.L. (2003). A stage is a stage is a stage: A direct comparison of two scoring systems. *Journal of Genetic Psychology, 164*(3), 335-364.
- Dawson, T.L. (2004). Assessing intellectual development: Three approaches, one sequence. *Journal of Adult Development, 11*(2), 71-85.
- Dawson, T.L. & Stein, Z. (2008). Cycles of research and application in education: Learning pathways for energy concepts. *Mind, Brain, and Education, 2*(2), 90-103.
- Dawson-Tunik, T.L. (2004). "A good education is....:" The development of evaluative thought across the life span. *Genetic, Social, and General Psychology Monographs, 130*(1), 4-112.
- Harris, L.S. and Kuhnert, K.W. (2008). Looking through the lens of leadership: A constructive development approach. *Leadership and Organization Development Journal, 29*(1), 47-67.
- Kegan, R. (1982). *The evolving self: Problem and process in human development*. Cambridge, MA: Harvard University Press.
- Kegan, R. (1994). *In over our heads: The mental demands of modern life*. Cambridge, MA: Harvard University Press.
- Kitchener, K.S. and Fischer, K.W. (1990). A skill approach to the development of reflective thinking. Developmental perspectives on teaching and learning thinking skills. In D. Kuhn, (Ed). *Developmental perspectives on teaching and learning thinking skills. Contributions to human development*, Vol. 21. (pp. 48-62). Basel, Switzerland: Karger.
- Lahey, L., Souvaine, E., Kegan, R., Goodman, R., & Felix, S. (1988). *A guide to the Subject-Object Interview: Its administration and interpretation*. Cambridge, MA: Harvard Graduate School of Education, Subject-Object Research Group.
- Lewis, P., Forsythe, G.B., Sweeney, P., Bartone, P., Bullis, C., Snook, S. (2005). Identity development during the college years: Findings from the West Point Longitudinal Study. *Journal of College Student Development, 46*(4), 357-373.
- Loevinger, J. (1979). Construct validity of the sentence completion test of ego development. *Applied Psychological Measurement, 3*(3), 281-311.
- Loevinger, J. (1993). Ego development: Questions of method and theory. *Psychological Inquiry, 4*(1), 56-63.

- Mascolo, M.F. and Fischer, K.W. (in press). The dynamic development of thinking, feeling and acting over the lifespan. In W. F. Overton (Ed.), *Biology, cognition and methods across the life-span. Volume 1 of the Handbook of life-span development*, Editor-in-chief: R. M. Lerner. Hoboken, NJ: Wiley.
- Stein, Z., & Heikkinen, K. (2008). On operationalizing aspects of altitude: An introduction to the Lectical™ Assessment System for integral researchers. *Journal of integral theory and practice*, 3(1), 105-138.
- Stein, Z., & Heikkinen, K. (2009). Models, metrics, and measurement in developmental psychology. *Integral Review*, (5)1, 4-24.

Katie Heikkinen is doctoral student at the Harvard Graduate School of Education, focusing on adult development—its assessment and its relationship to leadership outcomes. She is also on the faculty of the Integral Theory program at John F. Kennedy University, where she teaches multiple intelligences theory and cognitive development.

Email: katie.heikkinen@gmail.com