

STAGES background_Murray

Sentence completion assessments for ego development, meaning-making, and wisdom maturity, including STAGES

Appeared in Integral Leadership Review,[11](#) August 2017

Tom Murray
Open Way Solutions
tommurray.us@gmail.com

Table of Contents (clickable)

Overview	4
Background Context and Preface	5
An Introduction to Wisdom Skills and Adult Development	7
Why are wisdom skills important?	7
What is wisdom and how does it relate to meaning making?	8
Development: vertical vs. horizontal growth.	10
The shape of development.	11
Levels or stages of development.	11
What causes or supports development?	13
Caveats and Concerns—overgeneralizations about development.	16
Defining and Measuring Ego Development	17
What is ego development?	17
Ego development and the WUSCT.	18
On multiple intelligences and wide vs. narrow skills.	19
Cognitive vs. emotional skill—ego development includes both.	20
One skill to rule them all—is ego development one thing?	21
Cook-Greuter's study of Post-autonomous levels, and person-perspectives.	21
Total protocol scores.	23
Internal Validity of the SCT	26
Types of validity.	26
Inter-rater reliability of the WSUCT	27
Internal consistency of the WSUCT.	27
Test-retest reliability of the WSUCT.	28
Faking, guessing, and scaffolding the SCT.	29
Variations in the length of the SCT.	30
Variations in SCT stems.	31
Internal validity of WUSCT derivatives: MAP, GLP, and LDP.	32
External Validity of the SCT	33
Face Validity of the SCT.	33
Construct Validity of the SCT.	34
Construct and predictive validity of derivatives: MAP, LDP, and GLP.	36
On comparing models and construct indeterminacy.	37
The STAGES Model	38
What is STAGES?	38

Priorities of the four dimensions.	42
Further Explanation of the Four Dimensions	44
Concrete/Subtle/Metaware.	44
Individual/Collective.	45
Passive/Active.	46
Complexity: R/A/R/I.	48
STAGES vs. Prior Developmental Frameworks	48
STAGES vs. other SCT assessments.	48
Theory Comparison Caveats.	51
STAGES, States, and cognitive/brain theories	52
STAGES vs. Complexity and Skill Theory	54
Validity Studies of the STAGES Model	57
Conclusions	59
References	59
Appendix 1 — Summary of Conclusions in the Article	63
Appendix 2 — Developmental Level Correspondences	67
Appendix 3 — Descriptions of Developmental Levels	69
1. Short Summaries of Developmental progressions	69
2. Manners 2001, p. 544:	70
3. Loevinger's Ego Development Levels	71
4. Cook-Greuter (2004, p. 7)—Some examples of how different action logics matter	72
5. Table 1 from McCauley et al. 2006: Comparison of three constructive-developmental frameworks	74
6. Description of the STAGES levels and person-perspectives (PP), adapted from Fitch, 2016.	74

Overview

This article began as a series of short white papers providing various types of background information about STAGES and its predecessors.^[2] Its sections are relatively independent and readers with prior knowledge should be able to skip to and read each independently, or in any order. For those new to the field, it provides an overview of theories of meaning-making development (also called ego development or leadership maturity—and which we call "wisdom skills") as measured by the sentence completion test (SCT). For those familiar with the SCT, it provides (1) an overview of literature supporting its validity and properties; and (2) a deeper exploration of the meaning of meaning-making development. Finally we give an overview of the STAGES model.

STAGES is a new model of human development, created by Terri O'Fallon, that proposes an underlying system of factors driving the ego development and leadership maturity frameworks created by Jane Loevinger and modified by Suzanne Cook-Greuter and Bill Torbert. The STAGES model is relatively new and questions naturally arise regarding its validity and how it relates to other models. Thus, the purpose of this paper is to provide an in-depth look into background literature underlying the validity of STAGES.

The topics covered and questions addressed include:

- 1. Overview and Summary**—What are the themes covered in the article?
- 2. Background Context and Preface**—What is the larger context for this article?
- 3. An introduction to wisdom skills and adult development**—What is meant by the development of *meaning-making* or *ego development*? What are *wisdom skills*? Why are wisdom skills important? What is wisdom and how does it relate to meaning making? In developmental theory, what is vertical vs. horizontal growth? What are levels of development? What causes or supports development? What are some caveats and concerns one needs to keep in mind when using developmental theories?
- 4. Defining and measuring ego development**—We focus in on Loevinger's concept of ego development and its measurement through the WUSCT (as one way to measure wisdom skill). What is a projective test? What does the WUSCT look like and how is it used? Is "ego development" one thing or a combination of related skills or multiple intelligences? How does ego development coordinate the development of reasoning (cognitive) skills and social/emotional skills? How did Cook-Greuter's conceptions of person-perspectives and post-autonomous wisdom contribute to ego development theory?
- 5. Internal validity of SCTs**—What are the psychometric properties of the WUSCT and later SCT forms including MAP, LDP, and GLP? What has the research found about inter-rater reliability, internal consistency, and test-retest reliability? What types of modifications to the WUSCT are valid? What has research found about the vulnerability of the SCT to faking, guessing, and studying-for?
- 6. External Validity of the SCT**—How has the "face validity" or pragmatic usefulness of the SCT been argued for? How does the construct and measurement of ego development (meaning making maturity) relate to things like IQ, socio-economic status, psychological traits, and moral judgment? What does the research say about the ability of the SCT to predict or correlate with observable behaviors in leadership and personal growth and resilience?
- 7. The STAGES Model**—What is the STAGES model and how does it differ from prior models of adult ego development? How does STAGES describe development using four key dimensions drawn from Wilber's AQAL model? What is the relationship between stages and states in the STAGES model? How does STAGES relate to contemporary cognitive theories, brain theories, and Neo-Piagetian theories of hierarchical development? What conclusions can be drawn from STAGES empirical validity studies?

Appendix 1 contains a short summary of the arguments and conclusions contained in the article, which can serve as an "executive summary" for the reader.

In terms of both theory and empirical validation, the STAGES model rests upon theory and research on the MAP/GLP/LDP assessments of Cook-Greuter and Torbert; which in turn rests upon the large body of research on Loevinger's theory and instrument—the Washington University Sentence Completion Test (WUSCT). The quantity of research supporting the strength of the WUSCT and its ego development model is one or two orders of magnitude larger than that of the MAP/GLP/LDP and STAGES research combined. For example, a meta-analysis of over 350 empirical studies strongly supports the validity of the WUSCT assessment and the ego development model. Therefore, our argument for STAGES validity rests largely on WUSCT studies.

Background Context and Preface

In this preface I will explain what motivated me to write this paper—i.e. research and development in computer-based scoring—, which, other than in this Preface, is not a topic discussed in the rest of the paper.

The assessment of ego development or meaning-making development has traditionally been labor-intensive and thus expensive, requiring highly trained individuals to score text from essay question responses, sentence completions, or transcribed structured interviews. This limits doing large-sample-sized studies, and limits its general availability to larger audiences who could benefit. Beyond research

studies, the main practical application of these assessments has been in doing consulting and coaching work with professionals and small teams, where the cost of each assessment can be justified.

With recent progress in artificial intelligence, text analysis, and data science technologies it is possible to envision automated computer-based scoring of these assessments. In fact our group^[3] has developed the first such technology, called the StageLens technology, which can assess sentence completion tests with limited accuracy. It is not yet accurate enough to provide results for individuals, but it is accurate enough for aggregate results for statistical conclusions in sufficiently large groups (of 20 or more individuals). This opens the door to doing large-scale studies, including random samples of populations and whole-organization assessments.^[4]

Thus, the validity of the StageLens assessment rests in large part upon the validity of the STAGES model, which, in turn relies on the few validity studies of Cook-Greuter and Torbert's SCT variations, which, in turn rely heavily upon the large body of work validating Loevinger's WSUTC. Therefore, in this paper, in addition to giving readers an introduction to STAGES and ego/meaning-making development and measurement, we will step through the validity claims of each of these models.

Though our AI technology can be used to "learn" how to score based on examples from any large dataset of human-scored SCTs, StageLens has been trained with data scored using O'Fallon's STAGES model. The MAP, LDP, and GLP SCTs developed by Cook-Greuter, Torbert, and their associates are relatively minor variations on Loevinger's WUSCT test. O'Fallon's STAGES model is a more significant departure within this lineage. There have been fewer validation studies done with STAGES.

The intended audience of this article includes those interested in using SCTs for practical purposes, yet who want to know the research and validity background that supports the SCT. A main use of the SCT assessment is in helping individuals reflect upon their challenges, goals, learning, and psychological/spiritual growth, usually through coaching or consultant sessions. In such situations the psychometric properties and statistical validity of the assessments has been de-emphasized (though still considered important). This is understandable because the whole of a person's capacity cannot be reduced to a single number (a developmental level). Thus assessment results are usually used as a starting place for deep reflection, not for a definitive ranking or "high stakes" application. But as large-scale assessments become more feasible with automated scoring, attention once again turns to questions of psychometric validity and research methods.

This paper was originally drafted as a series of white papers to appear on the StageLens.com web site, to offer background information about the STAGES model and its predecessors to StageLens users. The occurrence of the STAGES Critique and Response articles in this issue of Integral Leadership Review has created an opportunity to publish that content more formally, as a published adjunct to the Response article. In the Response article ("A Response to Critiques of the STAGES Developmental Model") readers will find explanations of additional nuances about STAGES that are not covered in this article.

An Introduction to Wisdom Skills and Adult Development

Why are wisdom skills important?

A characteristic of modern culture is that we acknowledge that adults can psychologically change and grow over their lifespan—not just in storing new memories, learning new information, and being educated into new knowledge, but also in growing "developmentally" to change our most basic understandings of self, other, and world. It is a cliché that we live in a fast-changing, complex, and uncertain world, but we are only slowly coming to understand the capacities needed to craft wise decisions in the dynamic sea of diverse human activity and beliefs that we are thrown into in the 21st century. As Robert Kegan argues in *In Over Our Heads* (1994), the demands of contemporary society often outstrip the reasoning (and feeling) capacities of individuals making day-to-day decisions in home, civic, and work environments.

Three things become increasingly important in regards to these capacities—what we will also call "wisdom skills." The first is determining the *match* between the demands of a task, decision, or role and an individual's (or group's) wisdom capacities—i.e. are the skills well matched to the demands of the context?^[5] Second, assuming it is beneficial to do so, how do we *support and strengthen* any needed capacities? This article focuses on a third need that is a prerequisite to approaching the first two in realistic, practical, and systematic ways: we need to be able to *assess* these capacities to make good decisions about both matching and supporting them.

Our current research and development uses Terri O'Fallon's STAGES model of adult development (O'Fallon, 2011, 2013). STAGES is an extension of the work of Susanne Cook-Greuter on post-autonomous levels of development (1999, 2002); which is, in turn, an extension of the large body of work related to Jane Loevinger's model of "ego development" (later also called "leadership maturity" by Cook-Greuter and Torbert (2009)). In this introduction we aim to describe wisdom skill in terms of ego development and meaning-making development, and describe the basic principles of adult development for those new to the topic (including: vertical vs. horizontal growth, how levels transcend and include earlier levels, and what enables and supports development).

What is wisdom and how does it relate to meaning making?

Most of us have an intuition about the types of capacities that we think demonstrate wisdom or psychosocial maturity in adults. But how do we precisely describe this intuition about human potential—what is in common among those whom we admire as wise?^[6]

The focus of "21st Century" *education and workforce* development has been on skills such as: critical thinking; creativity, self-understanding, abstract reasoning, social/emotional/communication skills, curiosity and inquiry skills, understanding systems and wholes, and grounding ideas in pragmatic realities (AACU, 2007; NSTA, 2011; Clark et al., 2009; Pellegrino & Hilton, 2012; Scardamalia et al., 2012). A similar set of skills has been suggested for citizens to participate in a robust democracy (Muhlberger & Weber, 2006; Rosenberg, 2004, 2007).

Though these civic-participation and workforce-ready skills have a relatively pragmatic and cognitive focus, they overlap with the skills described in the sociological, philosophical, personality, and spiritual literature on "wisdom" per se (see Walsh, 2015; Baltes & Staudinger, 2000; Meeks & Jeste ,2009; Bangen et al. 2013; Fry & Wigglesworth, 2013). However, when we talk about the capacities that we expect in wise individuals (including leaders), additional skills come into play, including: multi-stakeholder perspective taking, robustness within paradox and uncertainty ("dialectical thinking"), empathy, humility and self-reflection. Wisdom skills (and meaning making maturity) involve the *application* of cognitive or reasoning capacities to the domains of human life (e.g. one's relationship to the beliefs and preferences of self and other). We can roughly summarize wisdom skills with these capacities:^[7]

1. *self-understanding, including* self-reflection, self-awareness;
2. seeing *big pictures*, including relational dynamics, contexts, and systems;
3. *perspective-taking*, empathy, compassion, and an appreciation for the diversity of human values, abilities, and contexts;
4. *tolerance of and appreciation for uncertainty*, paradox, and ambiguity;
5. a *humility* that includes being aware of the fallibilities in one's own beliefs and the limits of human reasoning in general; and
6. sound *pragmatic judgment* that accounts for observable reality, i.e. a balanced appreciation for the actual and specific complexities of human nature and of physical reality on earth.

When we think about the significant challenges of home, work, or civic life, it seems that such wisdom skills are both essential and too often in short supply. To understand, research, or take leadership around complex societal dilemmas the *assessment* of wisdom skills is important. Important enough to try, even though the difficulties in assessing (and even defining) wisdom will guarantee imperfect methodologies and models.

The above description of wisdom skills implicates a large and complex set of capacities. These include

cognitive (thinking) skills, affective (feeling) skills, personality traits, and values. However, to a first approximation, this set of skills can be successfully integrated into a single developmental construct, as has been done by Loevinger, Kegan, and others. These lines of research use what Loevinger calls "holist views of personality...[that] see behaviors in terms of meaning or purposes" (Loevinger 1970, p. 3). Kegan et al. (1998, p. 55) describe it as a "consistency in the structure, or order of complexity, of one's meaning-making (i.e. *how one thinks*)."¹⁰ These lines of research argue that the diverse set of wisdom capacities are closely related to each other, influence each other, and, in a *very rough* sense, tend to grow together as people mature (statistically and in general, though they don't necessarily track together with each individual). For example, Kegan et al. (p. 55) notes that Lahey (1986) found "an extraordinary degree of epistemological consistency within subjects across [the] domains" of intimate love relationships and more formal work relationships.

Robert Kegan and many others refer to this overarching capacity in terms of the complexity of one's "meaning making". Loevinger used the term "ego development." We also use the term "wisdom skills."¹⁰ By any definition these sub-skills, and the overarching construct, exist on a *developmental* continuum that ranges from more concrete, simplistic, black/white, right/wrong, us/them, win/lose modes of meaning making to increasingly more flexible, reflective, complex, and nuanced modes. The vast majority of scholarship on thinking skills, whether it be in the academic areas ("silos") of critical thinking, scientific reasoning, civic engagement, leadership and workforce readiness, wisdom studies, or personality traits, overlooks the developmental nature of these skills and does not cite the important research in adult developmental theory. Developmental models add key insights to the topic.

Adult development and learning is sometimes described in terms of general principles of change and growth without any measurable or specific levels,¹⁰ but here we focus on a wide terrain of models that do use a sequence of specific and assessable levels.¹⁰ It is beyond our scope here to describe the developmental levels proposed by Loevinger, Cook-Greuter, Kegan, and others, but we include some summaries in [Appendix 3](#) to give a flavor of how these levels are described in the field. Scanning those descriptions will give the reader a concrete understanding to ground the abstract properties of development we describe next.

Manners (2001, p. 549) describes Loevinger's theory as representing "an integration of diverse personality characteristics." Loevinger describes ego development as being concerned with "impulse control and character development, with interpersonal relations, and with [conscious] preoccupations, including self-concept" (Loevinger 1970, p. 3); and she describes the function of ego as a "structure of expectations," a "striving to master, to integrate, [and] to make sense of experience" (Loevinger, 1976, p. 59).

We can say that this holistic set of life-skills involves an increasingly mature and adequate understanding of the relationships among three realms: *self, others, and the world* (intrapersonal, interpersonal, and cognitive in Kegan's terms (1994); and "*I/we/it*" in terms of Wilber's Integral Theory (1995)).

Development: vertical vs. horizontal growth.

What do we mean by the "development" of these wisdom skills? Through experience, practice, and/or instruction people learn and grow their capacities, i.e. their skill and understanding. A rough distinction is made between horizontal and vertical learning. *Horizontal* learning involves learning more of the same thing—increased breadth, refinement, or differentiation of existing knowledge or skill.

Development refers to *vertical* (hierarchical) learning, in which a qualitatively new level of capacity emerges from the coordination, re-organization, or integration of a diverse set of lower level building blocks. A child may learn each of these skills separately: running, catching a ball, throwing a ball, and what it means to coordinate movement with others. To improve in each skill involves *horizontal* learning. To integrate them into the skill of "playing baseball" is a *vertical* transformation to another order of complexity and integration. Another example of vertical learning is the coordination of the adult skills of helping, self-reliance, leadership, etc., into the complex skill of "parenting."

The shape of development.

According to developmental theories, learning within any domain progresses roughly as follows: (1) in the early phases of learning one *struggles* to make sense of (or perform, or maintain a focused awareness of) some new type of thing (a concept, behavior, skill, idea, etc.); (2) once the capacity to deal with *individual* things of that type stabilizes, *horizontal* acquisition can proceed, and *more* of that type of thing is understood; (3) within horizontal growth one also learns about *relationships* between the things; and the type and complexity of these relationships increases; (4) at some level of breadth and complexity something rather magical happens, and a *new level* of simplicity and elegance emerges, often in a spurt of reorganization; (5) This reorganization creates a new thing at a *higher* (or later) level and the process starts all over again as one struggles to make sense of this new type of thing (Kegan, 1994; Commons et al. 1998; Fischer 1980).

Also, the early stages of learning usually begin in an external, factual, or "declarative" form—for example when a mentor describes or shows how to swing a golf club, how to solve a quadratic equation, how to drive a car, or how to parent well—and then one practices performing these instructions. With practice one moves from declarative to "procedural" knowledge, the ability to perform a task without needing to think about its steps. It is a movement from "knowing that" to "knowing how" (see theories of cognition such as Anderson, 1983; Laird et. al, 1987). Eventually, with enough practice, the skill becomes fully automated and effortless, and one may even forget the declarative knowledge that one started with (for example, one might develop deep skill in golf, driving, or parenting, but not be very good at explaining what or how they are performing it).

The levels of stage-like growth form an *invariant sequence*—levels can not be skipped, any more than one can learn calculus without first learning algebra, since each level forms structured wholes based on, or as Wilber puts it "transcending and including," the prior level(s). The term "simplicity" has been used to describe how an emergent higher level structure is both more *complex*, because it integrates lower level capacities and provides and adaptive response to more complex situations; and *simpler*, because of the transition from a complicated but chaotically organized skill set into a more elegant set of coordinated skills—an emergent skill at a new level.

Levels or stages of development.

Stage theories of adult development propose discrete stages or levels mapping from early to late stages of maturity.^[11] Theories differ on the number and granularity of levels, but a general concordance can be seen across all of them (Wilber, 2000; Stein & Heikkinen, 2009). Similar patterns of developmental progressions have been discovered in hundreds of studies across many areas, including ethics, values, perspective-taking, decision-making, epistemology (understanding the limits of knowledge), and spirituality (see Wilber, 2006). Most of these theories (beginning in the 1950's) described the general trend in terms of "pre-conventional to conventional to post-conventional" levels of adult development, and then define refinements within that spectrum resulting in between 5 and 12 levels. More recent work such as Kegan, Cook-Greuter, and O'Fallon's projects, have mapped out later stages, referred to as post-post-conventional, post-autonomous, trans-rational, post-formal, or dialectical (these are similar terms for the same territory, not a sequence of levels) (Pfaffenberger et al., 2011).

For example, Kegan's model has five stages of mental complexity (qualitatively different "orders of mind"). It frames development in terms of successive stages of becoming aware of aspects of the self that we were not previously aware of, yet they influenced our beliefs and actions. That is to look *at* what we previously (unknowingly) were looking *through*—which he calls turning *subject into object*. For example, we can learn to notice and describe our emotions in addition to simply having them; we can move from being able to plan to be able to think and talk *about* how to make a good plan. Kegan's (1994) *In Over Our Heads* eloquently charts how developmental maturation governs one's approach to the main domains of life: relationships, work, learning, parenting, and communicating.

Torbert and Cook-Greuter's model (based on Loevinger's work) includes nine "action logic" levels of ego

development or "leadership maturity," where each level involves a more sophisticated way to structure and coordinate thoughts and decisions (Cook-Greuter, 2002; Rooke & Torbert, 2005). O'Fallon's model, described later, includes 12 levels.

Cook-Greuter describes these action logics this way: "each meaning making system, world view, or stage is more comprehensive, more differentiated and more effective in dealing with the complexities of life than its predecessors...It describes increases in what we are aware of, or what we can pay attention to, and therefore what we can influence and integrate" (2004, p. 2,3). Thus, similar to Kegan's theory, what is being described is what one is not so much what one can *do* as what one is *aware* of, can speak and reflect on, and, eventually, can try to investigate, manipulate or change. It is useful to remember that one can perform a skill or ability automatically or unconsciously, but that does not "count" in meaning-making development until it "becomes object." For example, a child can *implicitly* make and follow a plan for how to get from their house, through the woods, to their friend's house. A developmental step is marked when they can talk *about* that plan and what is the best way to go (at the concrete level); and a further developmental level (called the formal, abstract, or subtle level) is marked when they can talk about the planning process, as in What is the most efficient way to make a plan, and How can we compare and combine our plans?

[Appendix 2](#) shows some correspondences between the levels defined in various theories. [Appendix 3](#) shows several descriptions of developmental progressions through stages. To describe each level or compare different models is beyond our scope, but the reader will appreciate the significant similarities in how they describe increasing complexity and nuance of I/We/it interrelationships.

These stage progressions have been used to describe development from childhood through late adulthood, to describe important meaning-making differences among adults, and differences in leadership style. In addition some theories use similar level descriptions but focus on developmental or evolution of cultures or groups (such as organizational cultures and anthropological epochs) (Wilber, 2006; Beck & Cowan, 1996; Thompson, 2007). The ability to use the same concepts describing increasing levels of complexity and integration (across I/we/it domains) to discuss both individual and collective capacity has been a powerful theoretical explanatory tool (though we must note that theories relating to cultures and groups have far less empirical validation, and sometimes the transfer of developmental principles from individual to collective may be more metaphoric than direct).

What causes or supports development?

Development occurs through a dynamic interactions between self and world. It is spurred when new contexts or information challenge a person's existing beliefs, skills, or frames of reference. To avoid the uncomfortable dissonance of conflicting ideas, the meaning-making drive strives to either modify the new information to agree with the existing frame of reference ("assimilation"), or modify the meaning-making frame to harmonize with the new information ("adaptation," and see Kegan, 1982). The greater complexity *and* greater simplicity of a new developmental level comes from its ability to find a higher perspective in which ideas (or skills) A and B, which seemed previously contradictory or disjointed, can meaningfully and productively co-exist.

It is widely noted that, as Kegan put it "people grow best where they continuously experience an ingenious blend of support and challenge" (1994, p. 42). The concept was originally articulated as the "zone of proximal development" by Vygotsky, 1978 and O'Fallon and Fitch speak of development spurned by "disorienting dilemmas." Across different life contexts an individual can display a range of developmental levels. King & Kitchener (2004, p. 9) state that "variability in individuals' responses across tasks reflects the degree of 'contextual support' available at the time (e.g., memory prompts, feedback, opportunity to practice)." Tasks that require performance without support elicit a person's "functional level" capacity, while tasks that provide contextual support can elicit performance at an "optimal level" that is closer to the upper limit of the person's cognitive capacity.[\[12\]](#)

Vertical development is both very gradual and not guaranteed, especially beyond conventional levels, as the cultural surround tends to resist ideation that threatens the status quo. It is generally thought that it takes 3-5 years to grow one level, if conditions are right. Loevinger (1979, p. 303) noted that "attempts

to raise ego level experimentally in a few weeks have not succeeded, but experiments that have lasted 6 to 9 months have had statistically significant success [on average a fraction of a level]. [And] people can lower their score more reliably and decisively than they can raise it." Cook-Greuter (1999, p. 52) says that we can "only conclude that an individual performs at least at such and such a stage under the test conditions, but one cannot exclude the possibility that they might operate at a higher level if given support... probed for further explanations and meaning...or given altered test instructions".

A detailed discussion of what is thought to support wisdom skill development is beyond our scope here, but processes hypothesized to support development include (and see Wilber et al., 2008):

1. *contemplative* and reflective practices,
2. social opportunities for *feedback* and reflection,
3. targeted deep psychological or psychoanalytic work, such as "*shadow work*,"
4. participating in social contexts where *diverse perspectives* are represented and discussed,
5. *associating* with individuals or community at later developmental levels, and
6. learning new *models or concepts* that embrace and integrate seemingly contradictory ideas.

Again, it is important to note that these are the things believed to support development, in part from anecdotal evidence but also from empirical studies about what works on average. But that there are many unknowns and contextual factors involved, and there is no evidence that these things will guarantee growth for any individual.

To continue reviewing the evidence, Vincent (2013) notes that "personality characteristics may enhance or constrain ego development" and in particular ego development has been found to correlate somewhat with Openness on the Big-5 Personality Inventory, and that "a preference for Intuition on the MBTI was associated with significantly higher ego development on program entry and with greater ego development during the programs" (p. 197).

Pfaffenberger (2005, p. 290) summarizes studies suggesting that the challenge of a difficult situation is not enough to spur growth, but that "accommodation [i.e. growth] was not related to the experience of difficulty per se but to seeing it as challenging one's worldview and to consciously struggling with the event...[and]...therapy may promote development exactly because the conscious engagement in life problems seems to be what facilitates growth and therapy often engages this kind of process." King (2011) describes evidence that development is not simply a result of dealing with life challenges, but that growth also requires an *attitude* of active and reflective engagement with those challenges.

Torbert (1994) describes a study of introducing an "action inquiry" approach to managerial training. It emphasizes deep reflection into the relationships between meaning-making and action within the I/we/it spheres. This led to notable increases in outcome developmental levels (though the sample size was small). Alexander et al. (1994) conducted a longitudinal study showing that graduates of a college emphasizing meditation and spiritual philosophy resulted in significant gains in ego development vs. comparative liberal arts and engineering colleges.

In a study by Vincent et al. (2015) evaluating an "enhanced" community leadership development (CLP) program that included additional psychosocial challenges such as experiences that are interpersonal, emotionally engaging, personally salient and structurally disequilibrating, found that for later conventional consciousness stages "enhanced CLPs were significantly more successful in triggering post-conventional development" (238).

Though the above is about the development of individuals, many authors have proposed similar developmental progressions for the evolution of culture (Wilber, 2007; Harari, 2014; Rifkin, 2009; Diamond, 1998; Thompson, 1998). Developmental advances in societies can lead not only to new understandings, but to new technologies and new social structures (and vice versa, in what Wilber calls "tetra-enaction"). Each later structure (social or cognitive), while solving some problems, will inevitably introduce new dilemmas. For example, the inventions of fire, shelter, money, and computers—all solved one set of problems while creating other problems. If conditions are right, over time this can create a spiral of ever-increasing development and complexity in culture and society.

Note that developmental progress is in no way guaranteed. People and cultures can respond to

dilemmas by ignoring what does not fit the current frame, or even regress to developmentally prior modes of response to avoid dissonance or perceived threat. Sometimes, in situations of threat or calamity, it is completely appropriate to "down shift" to earlier, more black-and-white or quick-and-dirty, modes of reason and action.

Caveats and Concerns—overgeneralizations about development.

We end this brief introduction to adult development with some caveats. Clearly the developmental perspective is powerful, both in its potential to help us make sense of the human condition, and (as we show later) in its scientifically validated ability to predict certain human characteristics. But it is important to note that the more powerful a theory is, the more susceptible it is to misunderstanding, over-generalization, and misuse. Fleshying out all of the caveats is beyond our scope here (see Stein, 2008; Murray, 2011), but here are some important caveats:

- (1) Though the models describe general patterns, it is important *not to pigeon-hole* individuals into caricatures—individual differences are as profound as the general trends. One's developmental level probably has *something* to say about how one approaches parenting, relationships, work, learning, etc., but does not predict exactly how any individual will think or act in any context.
- (2) Though we might speak of someone's "center of gravity" as their most common meaning-making level, people embody a range of levels at different times, depending on the challenges and supports present in any context.
- (3) "Higher" (later stage) development is *not always needed or useful*. If an individual is thriving adequately within their context, why push development? (In fact, doing so can be harmful.)
- (4) Though we focus on the broad holistic capacity of "meaning making" that generally coordinates over many developmental lines and life contexts, *each line or context* can develop at different rates. This not only argues against pigeon-holing (#1) but makes the point that moral/ethical or spiritual development can lag behind ego or cognitive development. One can develop very sophisticated meaning-making capacities relating self-other-and-world, and *still be malicious, narcissistic*, or otherwise socially deviant.^[13]
- (5) *Healthy horizontal* development at each level is (arguably) more important than vertical development. Each level builds upon prior levels, and weaknesses or pathologies can exist in foundational layers of the psyche. In other words, the "shadow work" of cleaning up earlier levels of development is usually more important to the overall health of self and the world than pushing for vertical development.^[14]

Sometimes the most important things in life are the most difficult to measure or assess, but that should not obstruct our inquiring. For example, NGOs and even some countries are increasingly interesting in measuring constructs like "gross national happiness," or environmental health—difficult but noble and important goals. Similarly, we will find that the development of wisdom or meaning making maturity is difficult to define and measure, but this difficulty has not prevented significant research into the issue. Likewise, the many caveats and limitations involved with putting the theory and measurement of "wisdom skills" to use should not deter our inquiry, but motivate a diligence in method and ethics.

Defining and Measuring Ego Development

What is ego development?

In the Introduction we described "wisdom skill(s)" and the related constructs of ego-development and meaning-making maturity to illustrate the importance and nature of these constructs for those new to

the topic. Here we will focus in on Loevinger's concept of ego development—exploring the definition and range of the construct from an academic, research, and validity perspective.

We can put the work of Loevinger and her successors into the larger context of constructive-developmental models. Jane Loevinger's (1966) theory of ego development and Robert Kegan's theory of meaning-making development (1994) have substantial overlap, as noted in prior sections. However, they use different measurement methods, with Loevinger using a "semi-projective" sentence completion test and Kegan using a structured interview method.^[15] Both are quite labor-intensive to score as compared with psychological tests that use multiple-choice or fixed-answer methods, and in return, the more complex methods are thought to yield deeper and more valid results than the self-rated fixed-choice methods. The interview method is particularly labor-intensive—both to take/administer, and to score. In discussing the *measurement* of meaning-making in this section we focus on the sentence completion tests (SCTs) of Loevinger and colleagues, while above we have also drawn conceptually from Kegan's theory. Newman et al. (1998, p. 985) noted that,

Ego development is one of the most comprehensive trait constructs in personality psychology. It has been described as a 'master trait' (Blasi, 1976; Loevinger, 1966) in that it serves as a schematic frame of reference providing a meaningful organization for numerous more specific personality traits. The detailed conception, as formulated by Jane Loevinger (1976), represents both a developmental characterology or scale of psychological maturation beginning in childhood and a major source of individual differences in adult personality organization.

Browning (1987, p 113) notes that the ego development "[postulates] a series of developmental stages that are assumed to form a hierarchical continuum and to occur in an invariant sequence...[describing a] person's customary organizing frame of reference, which involves, in the course of development, an increasingly complex synthesis of impulse control, conscious preoccupations, cognitive complexity, and interpersonal style."

The above descriptions are characteristic of Kegan's theory as well—next we move specifically to the SCT and the WUSCT. Below is a description of ego development levels that includes insights from Cook-Greuter and Torbert (more descriptions of levels can be found in [Appendix 2, 3](#)).

Table 1: Ego Development Stages
(from Cook-Greuter, 2004, p. 279)

Stage/Action Logic	Main focus	% adult pop. N=4510
Alchemist and above Deep processes and intersystemic evolution rules principles	Interplay of awareness, thought, action, and effects; transforming self and others	2.0
Strategist Most valuable principles rule relativism	Linking theory and principles with practice, dynamic systems interactions	4.9
Individualist Relativism rules single system logic	Self in relationship to system; interaction with system	11.3
Achiever System effectiveness rules craft logic	Delivery of results, effectiveness, goals, success within system	29.7
Expert Craft logic rules norms	Expertise, procedure and efficiency	36.5
Diplomat Norms rule needs	Socially expected behavior, approval	11.3
Opportunist and below Needs rule impulses	Own immediate needs, opportunities, self-protection	4.3

Ego development and the WUSCT.

Jane Loevinger developed and refined the Washington University Sentence Completion Test (WUSCT) over several decades as a tool to measure what she eventually called "ego development." The WUSCT is a 36-item sentence completion test. The items of the WUSCT address a variety of issues, including how respondents perceive and respond to personal relationships (e.g., 'My mother and I...'), authority (e.g., 'Rules are...'), frustration (e.g., 'If I can't get what I want...'), and everyday issues (e.g., 'Raising a family...'). Loevinger's approach was primarily data-driven, in that the theory or model of development grew out of analysis of large amounts of assessment data, in contrast to frameworks that start with a theory and develop assessments from there. In addition, scoring the assessment is exemplar-based—i.e. the scoring instructions contain thousands of categorized (actual) sentence completions, rather than descriptions of conceptual or syntactic properties of the completion text.

A distinctive aspect of the WUSCT is that it is a *projective* (or semi-projective) instrument, as is the Rorschach Inkblot test or any free-association task. This contrasts with other types of psychological and cognitive assessments including multiple-choice self-rating surveys (or sorting or peer-rating instruments), structured interview methods, reflective problem- or dilemma-solving tasks, and behavioral assessments. In projective tests subjects are not asked to produce a good or correct answer (the instruction are simply "complete the following sentences"). Rather, respondents freely "project" their frame of reference, world-view, assumptions, etc. into their answers. Though the Rorschach test has come under scrutiny, Lilienfeld et al. (2000, p. 56) note that the WUSCT "is arguably the most extensively validated projective technique."

On multiple intelligences and wide vs. narrow skills.

Next we address the complexity of the concept of "ego development" itself. Is it really one thing—a clear "developmental line," or is it a combination of clearly identifiable components? Is it too vague or general to be defined and measured? First, we discuss what it means to delineate a discrete type or "line" of human capacity.

Gardner developed the concept of "multiple intelligences" (1983) which Wilber refined in his description of separate "developmental lines," including cognitive, ego (self-sense), values, morals, needs, faith/spirituality, emotional, and kinesthetic lines (2006). Though this model serves well as a first pass in trying to integrate many threads of psychological and developmental research, when one tries to apply it, it quickly becomes apparent that these categories are massively overlapping with each other (e.g. spiritual, ego, and social-emotional skills have much overlap).

Kurt Fisher's Skill Theory (1980) provides important insights into this issue (as does Hierarchical Complexity Theory (HCT), a similar theory by Michael Commons, Commons et al., 1989). Fischer claims that skills develop in response to the demands of real life task situations: "the skill level that a person displays...cannot be considered independently of the context in which that skill is assessed" (Fischer & Farrar, 1987, p. 646). Athletic skill, parenting skill, arithmetic skill, and musical skill are largely independent because the *task situations* in which we learn and use these skills usually have little overlap. At a more fundamental level, very basic human skills and emotional drives such as those dealing with reproduction, eating, and territory, seem to operate fairly independently because the task situations or life-needs they address are relatively independent.

But complex human social *contexts* such as communication, decision making, parenting, leadership, and learning have massively overlapping domains and characteristics, so the skills developed to meet these needs *must* be expected to be equally *overlapping* and difficult to separate. So, even though Skill Theory (and HCT) is used to create precise definitions of human skills so that they can be studied independently, Skill Theory can also be used to explain why it is so problematic to separate complex human capacities like ego-development, parenting, communicating, leadership, or self-understanding into separate well-defined developmental "lines." And it supports the validity of using a wide and harder-to-define construct like ego-development or meaning-making maturity.

One can note a similar phenomena within the domain of cognitive research on "higher order thinking skills," where one sees a plethora of constructs being studied: metacognition, reflective judgment, scientific inquiry skills, critical thinking, problem solving, creativity, etc. Wisdom (mature meaning making) seems to be related to all of these in some way. Though there is great value in studying each of these in isolation, once again we face the fact that there is massive overlap in how they are defined and understood. Deanne Kuhn's significant body of research illustrates the deep interconnections among all of these skill sets, and she discusses the definitional and methodological conundrums of teasing them apart as separate skills (Kuhn et al, 1999, 2000, 2008).^[16] Skill Theory explains why they are difficult to separate (though, again, for some types of applications it is quite appropriate to do so).

Cognitive vs. emotional skill—ego development includes both.

The so-called "higher order skills" are primarily reasoning skills, while ego development and related adult developmental models clearly also imply emotional/social skills. The most prominent or general categorization of human capacities is reason (thinking or cognitive) vs. affect (emotional or socio-emotional). But even this basic categorization is extremely problematic, both conceptually, because all real-world reasoning includes affective aspects, and scientifically, as brain science has discovered the deep interconnectivity of the higher and lower brain centers (Immordino-Yang and Damasio, 2007; Goleman, 1995).

It is self-evident that people can have very strong intellectual skills while having deplorable socio-emotional skills. Brain science and evolutionary biology clearly indicate that (basic) emotional processes are distinct from reasoning processes, while, as mentioned, also showing that they strongly influence each other. Rather than frame intellectual vs. socio-emotional skills as disjoint sibling capacities, we have found it best to describe wisdom skill as involving the *application* of higher order intellectual skills (judgment, metacognition, critical thinking, etc.) to the *domains* of self and other (I, you, we, us, them, etc.). Intellectual skills by themselves are understood to apply to the domain of "it" or outside objects, but these skills can be applied to, not *some* idea, but *my/your/our/their* ideas (or needs, values, etc.). This helps explain the intuition that a certain level of complexity in the "cognitive line" in some way precedes or is a prerequisite for ego development (wisdom skill).^[17]

One skill to rule them all—is ego development one thing?

Kegan and Loevinger are among those researchers who see ego or meaning making as an overarching and unitary trait. According to Jespersen et al. (2013, p. 229):

Loevinger's life's work has been devoted to charting the course of ego development through a series of predictable, hierarchically organized stages from early childhood through adult life...[The ego], for Loevinger (1976), is a 'master trait' of personality...a holistic process, a striving for meaning and self-consistency over time...[it] involves many dimensions of personality development, such as motives for behavior, moral reasoning, and cognitive complexity, as well as ways of understanding oneself and others...[the ego is implicated] in activities such as impulse control, cognitive functioning, interpersonal relationship style, and conscious preoccupations...

Theorists disagree on whether ego is actually a single master trait or is composed of an interlocking set of sub-trait that act as one factor. This nuance is unimportant for our current purposes. Numerous studies have used statistical methods (including homogeneity, factor, cluster analyses) to show that the ego development construct measured by the WUSCT "loads on a single factor," i.e. appears to be mainly measuring one and only one construct (Westenberg et al. 2004a, p. 606).^[18] In sum, several lines of reasoning can be used to support the validity of ego development (or meaning-making or wisdom skill as we use the terms) as a valid unified construct.

Cook-Greuter's study of Post-autonomous levels, and person-perspectives.

Susanne Cook-Greuter extended Loevinger's work by more closely mapping out the terrain of the later level, 'post-autonomous' or 'post-formal,' stages (1999, 2002). Loevinger's theory was based on a wide variety of populations, including pregnant mothers, prisoners, college students, etc., and is strongest in its descriptions of the most common stages. Cook-Greuter was interested in the characteristics of later stages of maturity, which one might assume are more prevalent among professionals, leaders, the college-educated, and perhaps even spiritual seekers. In her ground-breaking dissertation work she evaluated thousands of WUSCT surveys collected in prior studies, and from that drew out and re-analyzed data on the later stages.

Cook-Greuter refined the definitions and scoring procedures for the later stages, and added an additional level to Loevinger's model. Following this she continued her research while making a move that Loevinger was critical of: using the sentence completion instrument to score individuals and give them *coaching or consultation* (as opposed to restricting its use to research applications). Cook-Greuter, along with Bill Torbert, refined the theory and method further over years of field experience. This included the opportunity to collect data on many more late-stage individuals. Modifications of the WUSCT were created (see Torbert 2014 for a comparison of Cook-Greuter's MAP, and variations called GLP and LDP, which all share 80% of the same stems with the WUSCT). Some sentence stems were changed to better fit the needs of business leadership or personal-growth contexts. Scoring manuals were extended and refined. The business of developmental scoring became a business, expanding the potential for the developmental perspective to benefit society, while also taking on the risks and concerns implied in commercializing the results of an academic project.

Though Loevinger was not explicitly trying to focus on any part of the developmental progression, she was, through constraint of the data gathered, or from implicit preferences, focused more on the conventional and early postconventional levels. Loevinger's theory focused on changes in impulse control, goal orientation, interpersonal relations, and conscious preoccupations, and these (particularly the first two) changes characterize growth from pre-conventional to conventional to early post-conventional stages. [Appendix 3](#) contains a number of comparable description systems for developmental levels that will give the reader a feel for them. Describing them in detail is beyond our scope, but in describing Cook-Greuter's contributions we will give a bit more detail about the later levels that she mapped out.

In later stages, as one gains some freedom from, or at least perspective on, one's own cultural conditioning, the focus of the transitions is different. Cook-Greuter (1999, p. 3):

Conventional stages describe forms of meaning making that seem required for adults to function in the roles of modern societies. Postconventional ego development, on the other hand, describes the rarer stages of meaning making in which some adults begin to deliberately and consciously wrestle with culturally programmed responses to life. They begin to examine previously taken-for-granted assumptions and explore the fundamental questions about knowing and reality.

Prior theories told a developmental story of *increasing* competence, autonomy, and social awareness with increasing development. Cook-Greuter did not refute this pattern, but discovered that for the later stages there is a *deconstructive* move as well, in which the limitations of knowledge and the ambiguity of the self-system become apparent. Her description of later stages of development includes:

The ego becomes transparent to itself; [one] looks at all experience fully in terms of change and evolution [and one becomes] aware of the ego's clever and vigilant machinations at self-preservation...[One becomes] cognizant of the pitfalls of the language habit [and starts] to realize the absurdity [or] limits of human map making....[one remains] aware of the pseudo-reality created by words...[and becomes] aware of the profound splits and paradoxes inherent in rational thought...Good and evil, life and death, beauty and ugliness may now appear as two sides of the same coin, as mutually necessitating and defining each other. (Cook-Greuter, 2000, p. 21-30).

Another key insight from Cook-Greuter was to frame developmental progression in terms of "*person*

perspectives." Scholars have noted how children's cognitive development increases in complexity by moving from first to second to third-person perspectives. The second person perspective is the ability to imagine or acknowledge the perspectives of others, as is required to move from narcissistic self-interest of the toddler into the adult world of social conventions. Third person perspective is the ability to imagine what *any* reasonable person would think, i.e. being able to reason "objectively," as is required in the modern world of scientific thought and democratic deliberation. Cook-Greuter discovered that this framework could be extended into fourth and fifth person perspectives (and, theoretically, further) as a unifying frame for ego development. This roughly parallels Kegan's hierarchical subject-to-object progression, as each level can "see" and think about the prior one as an object of reason.

Thus, though development through earlier action logics usher in changes in traits like impulse control and goal orientations, later stages are accompanied by changes including openness to ambiguity and dissonance. Barker & Torbert (2011, p. 55) report on Nicolaides (2008) study: "unlike people at conventional action logics who tend to try to avoid ambiguity, all of her post conventional sample saw positive, creative potential in ambiguity. But within this broad similarity, she found for distinctive responses to ambiguity: the Individualist, Stage 7, *endured* it; the Strategist, Stage 8, *tolerated* it; the Alchemists, Stage 9, *surrendered* to it, and the Ironist, Stage 10, *generated* it."

Total protocol scores.

Before moving to the section on the validity of the SCT, we should describe one technical detail about how the test is interpreted. In scoring a sentence completion protocol (or "inventory") each item is given a score, independently of the rest of the items, and then the 36 scores are summarized into an overall score. Loevinger and others are most interested in determining a "center of gravity" score, or Total Protocol Rating (TPR) and are less concerned about the nuances and contour of levels across all 36 items.^[19]

Deciding how to calculate this overall score turns out to be a complex question. One reason is that participants are expected to exhibit a range of developmental levels in their answers. In a projective test the subject is not trying to score as high as she can, and it is actually seen as more healthy or well-rounded when responses range over at least 3 or 4 levels (for individuals of low developmental levels there is less room to range over). Taking the *average* over 36 stems does not work. Our intuitive understanding of development tells us that if someone attains a score of, say, Strategist level, on 8 of the 36 items, then they must have a strategist level of meaning-making, because, according to both the theory and empirical evidence, it is difficult to "fake" or "guess" these items to produce a score higher than your actual level. Whether that individual has 5 or 15 scores at lower levels should not affect the center of gravity score. But having more low than mid-level scores *does* strongly affect an average over all stems.

The same limitation exists for taking a sum of the items, or using the mode. None of these methods match our intuition about the construct. (And, disconcerting as it may be, as explained elsewhere, in the end it is a shared *intuition* that grounds the meaning of the construct.) One way to compensate for this is using a weighted sum (or average) that gives higher scores more weight.

In fact, two separate methods are used to calculate overall protocol score. The *Total Weighted Score* (TWS) is just that, a sum over the items giving higher weights to higher scores. But the more often used method is the *ogive* method for arriving at a TPR, which assigns cutoffs for each level. The ogive formula is not a mathematical expression (as is the TWS) but a classification *procedure* that goes something like this: if there are at least 4 scores at level E or higher, score it at E; if there are at least 6 scores at level D or higher, score it at D... and so on from the highest score to the lowest (it is more complicated but this description suffices here). The ogive produces 8-12 (depending on the model) discrete levels and is thought to best match our intuitions about "levels of development." The TWS produces an integer value ranging usually from about 100-400, and this continuous metric is preferable for some types of research studies. Holt et al. describe Loevinger's reasoning this way: "Human development has this odd, psychometrically inconvenient property of maintaining the potentiality to respond on many lower levels after one has, in a certain sense, left them behind. The ogive rules respect this peculiarity and allow for it" (Holt et al., 1980, p. 917).

But how does one determine the exact cutoffs for the ogive method (or the weights for the TWS method)? For the ogive method Loevinger used probability theory (Bayes Theorem, Lee, 2012) to estimate how much evidence was needed to conclude that person was at a given level.^[20] Though various statistical and comparative methods can be used to support the validity of any set of choices, each method depends on an ad-hoc choice of parameters (such as false positive rates or weights) and in the end the choices are uncomfortably and unavoidably arbitrary and linked to intuition—and there is no single "correct" right way to determine the cutoffs (or weights). Arguments are made in the literature, described in later sections, that the overall method, including the cutoffs used, has excellent validity. Though we cannot say for sure that some other method of aggregating the 36 items might be just as valid.^[21]

Loevinger (1998, p. 5) describes the early process of defining the levels: "because we initially had no scoring manual, we discussed as a group how to classify each completion, *trying to imagine* the type of person who would give such a response" (emphasis mine). This lead to the first of a series of scoring manuals, all of them exemplar-based, i.e. a completion is scored by matching it to a set of real examples (which, in later manuals, are grouped into categories based on thematic similarity). To train to be a scorer is to learn to assimilate the intuitions embedded in the manual's exemplar organization.

As could be expected, differences between (trained) scorers tend to average out, and scorers were "more confident on judging the ego level of a total protocol than that of a single response out of context" (IBID, p. 5). Westenberg et al. (2004a, p. 485) notes "the scoring manual for the SCT [32-item youth version] consists of over 2000 response categories...about 80 response categories for each of the 32 items." The scoring manual used by Cook-Greuter, Torbert, and associates is over 300 pages long, with one chapter of examples for each of the 36 stems, with, usually, dozens of examples in each category. Training to use the manual takes many months and close supervision, until acceptable inter-rater reliability is achieved vs. other experts. Because human language is so diverse and expressive, i.e. because there are so many ways that an individual at level X could respond to stem Y, the creation of the example-based scoring manual requires the analysis of hundreds or even thousands of scored protocols.

Internal Validity of the SCT

Types of validity.

In this section we will summarize the literature on the validity of Loevinger's WUSCT and Cook-Greuter and Torbert's modifications to it (the MAP, LDP, and GLP).^[22] The literature drawing on Loevinger's model is so extensive that it includes a number of meta-analysis and critical overviews, substantially supporting its validity and usefulness (Cohn & Westenberg, 2004; Manners & Durkin, 2001; Holt, 1980; Novy & Francis, 1992; Jespersen et al. 2013; Westenberg et al., 200b).

For psychological assessments the "method" includes the data collection instrument, in our case the SCT, and the data interpretation, in our case the scoring method. The instrument includes the *instructions* given, and the scoring method includes the scoring *manual* and the method used to *train* the scorers. Changes to any of these can affect the validity of the overall method. In what follows when we speak of the validity of the WUSCT we are including all of these things, though we mainly focus on the SCT itself.

For our purposes, validity judgments exist in two broad categories: internal and external validity.^[23] *Internal* validity describes the quality, reliability, and repeatability of the assessment instrument or procedure itself, regardless of whether it is measuring anything useful, genuine, or meaningful. *External* validity describes how well a measurement instrument or experimental conclusion matches what it is supposed to measure or test. Are its results relevant and accurate across general real-world contexts—i.e. is it genuine and useful? For example, a weighing scale that is 10 lbs. too high has internal validity, in that you get consistent results from it, but it does not have external validity—it does not accurately measure what it is expected to measure.

Validity metrics are what compensate for the uncomfortable truth that psychological assessments are trying to measure something unobservable, vague and intuitive. An assessment with internal validity is a sound measurement of *something*, regardless whether it measures exactly what we intend. Strong external validity metrics support a claim that what is measured matches our intuitive or conceptual understanding of the construct.^[24]

First we will discuss *internal* validity in terms of: inter-rater reliability, internal consistency, and test-retest reliability; and then *external* validity, including face validity, construct, and predictive validity.

Inter-rater reliability of the WSUCT

Analyzing text or other qualitative data to derive a categorization or quantitative score is complex and uncertain business, usually requiring human judgment. Researchers compensate for the variability and subjectivity of human scoring by using multiple raters and measuring their agreement. A method is more objective and valid if raters tend to come to the same conclusion, and is less valid and too subjective if raters come to different conclusions. In the WUSCT literature several methods are used for assessing inter-rater reliability (IRR, agreement, or concordance), including percent agreement, Cohen's kappa value, and correlations (usually Pearson's R).^[25] Below we summarize those results without getting into the nuances of these different methods.

Westenberg et al. report that "psychometric studies of the WUSCT...invariably report high levels of interrater reliability. Perfect interrater agreement per item averages about 85% and interrater agreement within one stage (i.e., disagreement not larger than one stage) is often close to 95%" (2004a, p. 603). Cohen's Kappa values have been reported at about .80, which is considered excellent.^[26]

Those statistics were for agreement at the level of the total protocol rating. For agreement at the level of each stem completion, Loevinger & Wesler (1970, p. 41) report agreements averaging 77% (ranging 60% to 86% over the stems); and correlations averaging .75. Pfaffenberger (2011, p. 11) says the literature generally points to higher IRRs (near .90). Newman et al. (1998) report a per item weighted kappa statistic averaging .73 (ranging .47 to .93).^[27]

Internal consistency of the WSUCT.

Internal consistency measures the correlation between items on a test. It should be fairly high to indicate that all of the items are measuring essentially the same thing. However, for a test like the SCT, it is not expected to be too high because each sentence represents a different contextual perspective, and we expect individual differences in which contexts will reveal evidence about each person's highest capacity (like triangulating a measurement from different angles). According to Westenberg et al. (2004b, p. 693) "The WUSCT [displays] high internal consistency: Most studies report a Cronbach's alpha of .90 or higher" (e.g., see Loevinger, 1998; Novy & Francis, 1992; Minard, 2000; Newman et al., 1998).^[28]

Test-retest reliability of the WSUCT.

In general reliability refers to the extent to which repeated measurements yield consistent results. Test-retest reliability refers to whether an assessment measures a stable construct or something that varies due to uncontrollable or random factors. It also measures how taking a test again, by itself, influences the outcome. Meaning-making development is expected to change very slowly (some have estimated at least three years per level when conditions are supportive), though theoretically it can change fairly quickly when it does change, if the growth is in spurts.

The SCT assesses capacity at a moment in time, and it is possible that an individual is not "on top of their game" on that particular day and time.^[29] Also, as Loevinger puts it: "frequent measurement is likely to be resented and hence to result in poor validity for retests" (1979, p. 287).^[30] It is also possible that people can actually regress in their ego development, especially under periods of distress. Loevinger notes that "people can lower their score more reliably and decisively than they can raise it" (p. 303).

True test-retest assessments of the SCT, i.e. test of the stability and repeatability of results over short periods of time, are rare in the literature (perhaps because they are inconvenient for the test taker and time-consuming for the researcher) and most "retest" situations are longitudinal studies looking for growth over time. However, Westenberg et. al. (2004b, p. 603) report that the "test-retest reliability of the WUSCT [is] high, and test-retest correlations are often about .80." Manners & Durkin (2011, p 545) say "in terms of test-retest reliability, when sufficient time is allowed between the two tests to allow for motivational effects, significant correlations have been found between test and retest scores."

Faking, guessing, and scaffolding the SCT.

A related issue is whether the SCT can be faked, gamed, or studied for. The anecdotal lore in the practice community is that it is quite difficult for a person to score much higher than their "actual" capacity, which is the case with any valid assessment of a skill that grows hierarchically (e.g. one normally can't pretend to play the violin at a higher level than they are actually at). However, because SCTs are mediated by linguistic skill, one must have at least an adequate ability to articulate what one thinks or believes to have the score reflect their developmental level. Also, one can theoretically do better on the test by knowing how the test is scored and adjusting one's language accordingly. This is one reason why scoring manuals are often kept confidential (and those using SCTs in for-profit coaching and consulting ventures of course have additional incentives to keep scoring methods proprietary).

As mentioned above, psychologists differentiate between functional (or characteristic) vs. optimal (ideal) performance, where the former represents everyday or average cognition and the latter represents the maximum capacity possible within ideal or well-supported contexts. Projective assessments like the SCT, in which the subject is not asked to give a "correct" answer, tend to elicit functional performance, but also can show a relatively wide range of levels over the items. Structured interview assessments, where interviewers probe for deeper reasoning, tend to elicit more optimal behavior (Kegan et al., 1998). Problem-solving or dilemma-reflection activities elicit behavior that is in between those extremes, since participants are presumably trying to do their best at something. Multiple-choice assessments often expected to rate later than the other types, since they rely on recognition rather than recall or construction of knowledge (and self-rating fixed-choice questions can be biased toward higher results as well).^[31]

The instructions and context of the SCT can influence the outcome, especially if they provide any sort of support, "scaffolding," or prompting. Westenberg et al. (2004) studied the SCT's sensitivity to changes in the administration of the instrument. They found no significant difference between oral vs. written versions, and a very slight decrease in scores when it was administered orally via a telephone conversation. On changing the SCT instructions from simply "please complete the following sentences" to include instructions like "be candid" or "make a good impression," they conclude that "several studies suggest that such instructions do not appear to influence ego level ratings" (p. 694). In contrast, "three other types of instructions, each with conceptual relevance to ego level, had a modest but significant impact" (IBID). The additional instructions included things like answering "...in the most complex and thought-provoking way," and "...in as adult and mature a manner [as you can]..." In another study subjects were "provided with brief descriptions of each ego stage and instructed to complete the sentence stems as they would be completed by a person at the highest ego levels." For all of these methods Westenberg et al. conclude that the "average increase was no more than one half a stage."

All of the above suggests that the SCT can be influenced by various factors, but the variations caused by these factors are not more than a half a stage. However, even though there was a study where subjects were given a description of developmental levels, there is no study we are aware of that addresses how much a deeper instruction about development theory influences scores. This is an important question since many integrally-informed or developmentally-informed education/training/personal-growth

programs employ developmental assessments, and use them to test the hypothesis that the program or intervention supports adult development. Loevinger (2011, p. 70) notes that contemporary subjects "are more sophisticated and educated in developmental theory than prior generations...[such that]..trying to prove to oneself that one is at a later stage is another hazard [in scoring, and] to distinguish between genuinely mature integrated protocols and those that consciously or unconsciously try to game the test has become a feeling aspect of training certified scores."

More research is needed to know whether an observed change is truly from a deep transformation in meaning-making development or from a mere intellectual understanding and inculcation into a community valuing certain linguistic markers (or indeed, how to define the difference between those two things).

Variations in the length of the SCT.

We have mentioned that there are various versions of the WUSCT and its successors. These variations involve new stem choices and/or using less than the standard 36 number of items. Some studies use a "split-half" version of the test, giving 18 of the stems as a pre-test and the other 18 as a post test or alternate test. Some variations of the SCT are targeted toward men, women, or youth. Our goal here is not to compare these variations in any detail, but to argue for the robustness of the SCT over such modifications.

As to the *length* of the SCT: Novy & Francis (1992) compared the split-half versions of the WUSCT (18 items each). They conclude "these results provide empirical justification for those users of the SCT who have the need for shorter, interchangeable, and reliable forms of the test."^[32] Holt (1980, p. 909), experimenting with a 12-item short form of the WUSCT found that inter-rate reliability was "at least as good as...reported by Loevinger; and the internal consistency...was quite adequate....Analyses of other data indicate that the short forms are representative samples of the full [WUSCT]".^[33] He goes on to say the data show that "an abbreviated form of Loevinger's WUSCT is a reasonably reliable, feasible, and useful instrument for large-scale research...[and is]... a representative sample of the larger instrument, which probably gives substantially the same results" (p. 916), "...though it is clearly less satisfactory than the full 36-item form" (p. 915). Basic psychometric theory predicts that more evidence will result in better accuracy, so for individually-based assessments, done for coaching or consulting purposes, the full set of items is still recommended, but for research or group-statistical assessments, it would appear that shorter forms are quite valid.^[34]

Variations in SCT stems.

As to variations in the choice of sentence stems, we can make several observations. Loevinger adapted the sentence items numerous times before settling on a final version. There is nothing particularly special about the stems Loevinger used. Though they were carefully chosen and vetted; many were drawn from the experience of prior researchers, and the entire set evolved over the years before settling into the standard form used today. Others have used alternate forms tailored to men, women, or youth. Proposed new stems can be inadequate for a variety of reasons, e.g. they may be vague and thus understood in very different ways; they may coerce an overly limited range of responses; or they may introduce biases confounded with what the SCT is meant to measure. So new stems must be pilot tested for clarity and psychometric validity. But assuming that this due-diligence work is done to ensure that questions are adequate, overall the research strongly suggests that the overall properties of the SCT method, its strong psychometric properties, are robust to changes in the choice of sentence stems.

Ego development (meaning-making complexity) is a holistic capacity spanning all life-contexts (though we may exhibit maturity different than our "center of gravity" in any given context). The stems are meant to probe across a variety of contexts, and triangulate toward an overall measurement. Cook-Greuter (1999, p. 52) notes that "As Fischer, Hand, & Russell (1984) pointed out people tend to respond optimally to a task in 'domains in which they are highly motivated.' To tap this motivation, the SCT stems

were devised to address ordinary everyday experiences shared across a wide spectrum of people."

We have mentioned Kegan's and Wilber's framing of a holistic span of life contexts as including subjective, intersubjective, and objective (I/we/it) contexts. Loevinger (1985, p. 424) describes the span of stems in a related way: "Looking at item content, the stems can be classed as *first person* (My father—, When they talked about sex, I—), *third person* (Sometimes she wished that—, Usually he felt that sex—), and common noun or *impersonal* (A good father—, Being with other people—)."

Some stems seem to be sensitive to particular transitions along the developmental spectrum, and this is another reason for having an adequately diverse set of stems. For example, some stems are more related to impulse control (though one can give evidence about impulse control in any stem). Impulse control comes on line abruptly at conventional levels (second person perspective), and increases gradually or levels off for later levels. Therefore, sentences sensitive to impulse control are also sensitive to the transition from pre-conventional to conventional levels. Abstract and formal thinking begins at third person perspective with a qualitative leap, and increases gradually or levels off after that, so we would expect that certain stems are more sensitive to transitions in this part of the spectrum. So in general we would expect that, to a weak but statistically significant degree, certain sentence stems are better at signaling changes in specific levels (we are not aware of any detailed research results on this point).

Torbert and Cook-Greuter modified the stems to "omit a number of gender-based items and, includes work or leadership-related stems" (Torbert & Livne-Tarandach, 2009). Torbert's LDP is shorter (24 items) and includes six new stems not in the WSUTC. Torbert (IBID, p. 134) reports that "the responses to the new stems correlate better with an individual's overall profile rating than responses to the former stems did, thus improving the overall reliability of the measure." Later we will discuss changes made in O'Fallon's STAGES model.

Internal validity of WUSCT derivatives: MAP, GLP, and LDP.

As mentioned, Cook-Greuter's, Torbert's, and O'Fallon's works builds upon Loevinger's and branch off in several ways. First, all have made minor modifications to the set of sentence stems. Second they have each repurposed Loevinger's research-only methods for use in commercial coaching, consulting, and assessment ventures. Third, they include elaborated descriptions and scoring for the later stages, and they draw on populations consisting of more professionals or highly educated individuals. (O'Fallon diverges even more, as explained later.)

We have mentioned the implications of the first (stem modifications), and have also noted the second (commercialization). As to the third (later stages): theory predicts, and research indicates, that, because each level builds upon prior levels, that there are more degrees of freedom, diversity, and individuality in mean-making as stages progress. Later (post-conventional) levels have more linguistic variability in themes, styles, and syntactic complexity (though at the highest levels, Construct Aware and above, sentence completions tend to become shorter again).

Also, later levels are rarer than middle levels, making it more difficult for a research team or community of practice to consolidate upon the definition. In addition, above Construct Aware the scoring manual descriptions are less detailed and more vague. Because of these factors one can expect that the later the stage the more difficulty it may be to score (or to precisely define scoring criteria). Indeed trained scorers consider later levels more difficult to score (by anecdotal evidence). The diversity of responses and difficulty in scoring in part explains the fact that Cook-Greuter's and Torbert's studies tend to show lower internal validity vs. Loevinger's. The statistics for these adapted instruments are still quite acceptable, however.

Torbert & Livne-Tarandach (2009) report on variations of the WUSCT that have evolved into Cook-Greuter's MAP and Torbert's LDP and GLP instruments. They report an IRR of 69% perfect matches and .90 within one level, and a Cronbach's alpha measure of internal consistency of .91. Compare this with that above findings of the WUSCT: Perfect interrater agreement per item averaging 85% and interrater agreement within one stage often close to 95% (Tobert & Livne-Tarandach used alpha metrics while others have used kappa values, which can not be easily compared). Torbert (2014, p. 7) reports an inter-

rater study of "805 measures, each of which could have been scored at 13 different levels [using early and late- specifications of the developmental levels]. The result showed a .96 Pearson correlation between the two scorers, with perfect agreement in 72% of the cases, with a 1/3 action-logic disagreement in 22% of the cases, and with only one case of a disagreement larger than one full action-logic."

In sum, variations on the original WUSCT continue to have excellent internal validity metrics, showing the SCT method is sound and extensible for measuring meaning making. However, the internal validity metrics of the newer versions tend to be a bit lower, presumably because the populations being studied have a higher developmental level on average—later levels have more variation in responses, and are more difficult to score (and more difficult to describe in scoring manuals).

External Validity of the SCT

External validity describes how well the SCT matches what it is supposed to measure or test (ego development, making-making development, or wisdom skill). Is the test relevant, accurate across real-world contexts, genuine and useful? We will look at face validity, construct validity, and predictive validity.

Face Validity of the SCT.

The term "face validity" is used to summarize a qualitative or intuitive, i.e. not psychometric, argument for the effectiveness and usefulness of a measurement or procedure. Phaffenberger (2011, p. 10) says that "the face validity of the SCT is demonstrated by the sheer fact that it has been used in more than 300 research studies [including] such diverse topics as parenting behaviors, managerial effectiveness, and the effects of meditation on recidivism rates." Blumentritt (2011, p. 153) say that "more than 1,000 articles and book chapters have been published examining nearly every conceivable aspect of the construct and measurement of ego development," overall showing "substantial support" for the theory and measurement.

The face validity of the MAP/LDP "leadership maturity" instruments is further evidenced by their growing popularity in certain professional and leadership development circles. Additional face validity comes from the fact that "the WUSCT has been translated into 11 different languages in 19 different published studies" as of 2008 (Minard, 2009).

Anecdotally, it is also widely reported that individuals who take the assessment and receive the associated personal debriefing believe that the process "gets them," that something of their essence and life situation is well captured, and that the process helps them reflect on strengths, weaknesses, and strategies in pursuing life goals. Torbert (2014, p. 8) reports that over about 800 hundreds of clients who were asked to comment on the results of their SCT (GLP) score, "clients' estimates differ from the GLP in only 9% of the cases. In 2/3 of the cases of disagreement, discussion has led to client agreement with the GLP score; in the other 1/3 of the cases the GLP debriefer came to agree with the participants' self-estimate."

The face validity of the instrument is also supported through the experience of scorers. Cook-Greuter (2011, p. 66) says: "those who are skilled in interpreting the SCT are continuously fascinated with how much one can deduce—and intuit—about a person's experience of life from just those 36 sentences."

Construct Validity of the SCT.

Does the thing we are actually measuring with the test have the expected relationships and correlations with related phenomena? According to an overview by Westenberg et al. (2004a) the WUSCT has very

strong psychometric properties, having "indicated excellent reliability, construct validity, and clinical utility" (p. 596). As of their 2004 article, "findings of over 350 empirical studies generally support critical assumptions underlying the ego development construct" (p. 485), and dozens more studies have followed since 2004.

The construct validity of a psychological measurement is determined through its correlation with other known measurements. To demonstrate that the test measures what it is supposed to measure it should have *some* correlation with other constructs that it is *expected* to correlate with. However, to demonstrate the *usefulness* of a construct or test, it should not correlate *too* closely with sibling constructs, because if it does it is probably replicating another one, and is not very useful (though it *is* useful if it is easier to administer, score, etc.). In technical terms, construct validity looks at four things: (1) *concurrent* validity refers to whether the test corresponds to previously established measurements for the same construct; (2) *convergent* validity refers to its correlation (association) with other measures that it is theoretically expected to relate to; (3) *discriminant* validity refers to how non-correlated it is with other measures that theory says it should not be related to; and (4) *incremental* validity refers to its ability to account for variance (or increase predictive power) over and above that explained by other measures. The research summarized below explicitly or implicitly addresses each of these—note again that the studies span some variation in the SCT design (e.g. for youth, men, or women; or of short vs. long forms).

Numerous studies have shown that ego development is distinct from but correlated with *intelligence* (about a 30% correlation) and *verbal fluency*, and, as expected, like (IQ tests), it correlates with *socioeconomic status*.^[35] Pfaffenberger says "the SCT correlates with *education*, socioeconomic status, and *complexity of work*, which has been shown to hold true across international samples. This is not surprising because education and social class relate to aspects of impulse control, goal orientation, and conscious preoccupations, which is exactly what the SCT is meant to assess" (2011, p. 12; emphasis mine). Pfaffenberger reports that Loevinger (1998) found that the WUSCT correlates with *verbosity* (i.e. the length of sentence completions) at about .31. She notes that despite its correlation with other measurements, "the SCT does not intend to predict behavior, measure social adjustment, or evaluate psychopathology" (p. 10).

In their meta-research Westenberg et al. (2004a) summarize construct validity findings as follows (italics is mine):

(a) longitudinal studies have confirmed the invariance of the developmental steps (i.e., *no stage can be skipped*; see Loevinger, 1998); (b) average increases with *age*...have been documented (Cohn, 1998); (c) ego level maturity is related to a wide variety of relevant *individual differences* within age groups (see Manners & Durkin, 2001; Westenberg, Blasi, & Cohn, 1998); (d) ego level scores display *incremental validity over IQ and SES* (e.g., Browning, 1987; Cohn & Westenberg, 2004); and (e) the construct and the measure have proved applicable in different *countries, cultures, and languages* (see Carlson & Westenberg, 1998).

In her meta-research Lilienfeld et al. (2000, p. 56) summarize the WUSCT's construct and incremental validity as follows (italics is mine):

[The WUSCT has] impressive construct validity in numerous studies by independent investigators...and fulfills our criteria for empirical support. For example, scores on this instrument correlate (a) moderately to highly with ego level as assessed by *interviews* (e.g., Lucas, 1971), (b) moderately with scores on Kohlberg's (1981) *moral judgment* test even after controlling statistically for age (e.g., Lambert, 1972), (c) negatively and substantially with indexes of *delinquency and antisocial behavior* (Frank & Quinlan, 1976), (d) positively with successful *adaptation after divorce* (Bursik, 1991), (e) positively with the *openness to experience* dimension of the "Big Five" personality taxonomy (McCrae & Costa, 1980), and (f) positively with observer ratings of *ego resiliency and morality* (Westenberg & Block, 1993).

In addition, the WUSCT has demonstrated substantial incremental validity above and beyond intelligence measures in the prediction of personality traits among nonclinical participants (Westenberg & Block, 1993) and length of stay and problematic ward behavior among *psychiatric inpatients* (Browning, 1986). In a sample of *twins* reared apart, Newman and Bouchard (1998) also found that WUSCT possesses considerable genetic variance even after controlling statistically for the effects of intelligence measures. Finally, as predicted by Loevinger's model of

ego development, WUSCT scores have shown curvilinear relations with *measures of conformity* (Hoppe & Loevinger, 1977; Westenberg & Block, 1993).

Cohn (1998, p. 141) summarizes studies using peer mentoring or counseling experiences, what are thought to increase perspective-taking. "At least eight studies (mainly composed of high school students) have reported significant advances in WUSCT scores following such interventions, although participants did not, on average, advance beyond a Self-Aware stage."

In conclusion, the SCT is a valid measure of ego development, or meaning-making capacity as Loevinger defined it, as demonstrated by numerous indications of its construct validity.

Construct and predictive validity of derivatives: MAP, LDP, and GLP.

Above we summarized the internal validity of the MAP, LDP, and GLP variations of the WUSCT instrument, and some results of studies using them. Little has been done formally with these instruments to add to the formal construct validity studies done for the WUSCT. However, there seems to have been more in-situ/in-vivo studies involving these variations of the SCT than the WUSCT. These studies further demonstrate the face validity of the SCT and also contribute much of what is known about predictive or causative aspects of ego development and its support. (However, overall, such studies are difficult to control and scale, and tend to have had low participant numbers—so more research is needed.)

Bill Torbert and his associates have done the most research into the real-world causes and effects of ego-development (i.e. leadership maturity) and the effects of interventions designed to support ego development. Torbert (2014, p. 3) summarizes some of these studies: "we have found very powerful correlations (accounting for more than half the variance in the outcomes) in terms of the organizational action-logic necessary before an organization systematically supports leadership development at work. Another finding that accounts for more than half the variance shows which leadership action-logics are necessary to reliably succeed in generating organizational transformation. More specifically, only those CEOs and lead consultants who have measured at the Transforming or early Alchemical action-logics reliably generate positive organizational transformation (leading to larger market share, profits, and reputation)."

Vincent (2015) notes that "Although research in this areas is still in its infancy, a growing body of studies is showing associations between increasing consciousness development and better leadership performance and organizational outcomes"—and she cites a substantial 21 articles in this regard (p. 2). Literature overviews of this research usually include the frameworks of Kegan and Torbert/Cook-Greuter as measuring comparable aspects of leadership maturity.

Additional descriptions of additional studies of ego development within the leadership and management field can be found in McCauly et al., (2006). They found that "Dependent leaders [Diplomat, Expert] were reluctant to delegate, tended to avoid holding others accountable, sought unanimous agreement in their groups, felt threatened by others' complaints, had difficulty expressing their disagreement, and saw college authorities exclusively as judges and experts. In contrast, Independent leaders [Achiever and above] negotiated performance standards with group members and held them accountable, evaluated the complaints of others, expressed their disagreements with others, and saw college administrators as resources" (p. 240). The studies also suggest that later stage managers may be more likely to empower and inspire subordinates...leaders need to be at least at the Independent order to be effective transformational leaders." Included in their overview of many studies is a study of 41 executives by Harris (2005), who "examined the relationship between the executives' order of development and the ratings they received on a 360-degree feedback instrument. Order of development predicted the average ratings (across all raters) executives received on seven of the eight dimensions assessed by the instrument" (p. 640).

On comparing models and construct indeterminacy.

In the human sciences much attention is placed on the definition of the constructs used, and scholars often take what we could call a "construct aware" view that these concepts, such as intelligence, ego maturity, introversion, etc. do not so much exist (as defined) in the world as much as they are invented by people—and their meaning may both vary among scholars and communities of practice, and shift over time. In the end what a psychological construct means is tied up with how it is measured and the validity of those measurements is disconcertingly tied to human *intuitions* about the thing to be measured (Stein & Heikkinen, 2009; Lakatos, 1976).^[36] Each scholar must be as clear as they can about what their constructs mean, contrast their meaning with others', and argue for how their way of measuring the construct matches its intuitive meaning.

This fact becomes important when we consider variations on the WUSCT such as those of Cook-Greuter, Torbert, and O'Fallon (or even comparing these to Kegan's model).^[37] With each modification of the method one is measuring a slightly different contour within the complex phenomena that is human meaning making. Comparative analysis has shown that, within statistical tolerances, they are measuring *more or less* the same thing, but in the end there is no ground truth upon which to judge which is "closer" to the "real" phenomena of meaning-making (or ego development), or the definition of any particular level. It is the measurement method itself that "puts a stake in the ground," to allow for stability, standardization, and easy comparison—but that stake may need to move or fork from time to time as the field evolves its understanding, and as sub-projects with different perspectives branch out.

The STAGES Model

What is STAGES?

The MAP, LDP, and GLP SCTs developed by Cook-Greuter, Torbert, and their associates are relatively minor variations on Loevinger's WUSCT test. Some stems were changed, and the number of stems may be slightly different, but the instrument, the scoring method, and the underlying theory are essentially the same. Scoring is based on matching a subject's completions to categories of examples in a lengthy (over 300 page) scoring manual that has separate sections showing exemplars for each of nine developmental levels, and for each of the (typically) 36 sentence stems in a protocol (totaling about 16,000 examples organized into over 200 categories). O'Fallon's STAGES model is a more significant departure within this lineage.

Loevinger was clear that her theory or model was data-driven, in that the levels (and the scoring examples) were defined iteratively as researchers discovered and refined patterns intuited across large numbers of protocols. Cook-Greuter and Torbert followed suit in endorsing bottom-up model building, where theory emerges from raw data. The value being that biases introduced by preconceived theoretical notions of human development are minimized. These researchers, like all modern scholars, acknowledge that theories, explicit and implicit, and other preconceptions, unavoidably influence how data is interpreted, but they tried their best to let the data speak for itself.

O'Fallon's model is more heavily theory-motivated. She studied and scored within Cook-Greuter's system for many years, while also immersing herself in Ken Wilber's Integral Theory, which includes a "four quadrant, eight zone" theory of the fundamental ontological categories (primordial dimensions or ontological polarities) of reality (the AQAL model, see Figure 1). Wilber uses AQAL to theorize about the evolutionary/developmental processes underlying matter, biology, and culture (Wilber, 2000, 2006). O'Fallon's intuitions about human development were also informed by a lifetime working in various educational and service professions (including being a school principal). Through what can best be described as a stroke of intuitive pattern-matching, O'Fallon saw that the progression of developmental levels described by Cook-Greuter (and Loevinger before her) could be explained at a deeper level through the lens of the AQAL model. Over succeeding years and many conversations with Wilber and other colleagues, she refined the STAGES model into its current form. We will give a very brief

description of STAGES here, and refer the reader to item 6 "Description of the STAGES levels..." in [Appendix 2](#) (from Fitch, 2016), and (O'Fallon, 2013, 2015) for deeper explanations. The levels are illustrated in Figure 2.

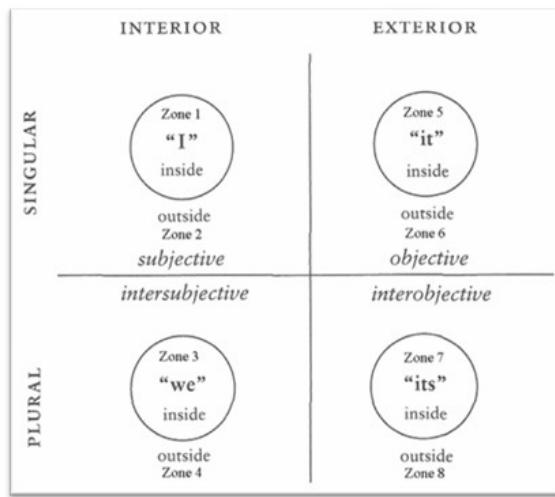


Figure 1: AQAL -Quadrant/8-Zone model, Adapted from Wilber, 2006

Wilber developed his AQAL model through an interdisciplinary meta-theoretic analysis of over 100 separate theories, many of them describing processes of complex change, evolution, or development (Wilber, 2000, 2006). The model proposes a small set of dimensions that can be used to coordinate both the objects of reality and the methodologies (or epistemological perspectives) used across all disciplines.^[38] AQAL describes both objects (things) and perspectives (ways of perceiving or understanding things) in terms of three primordial dimensions, each of which contains two categories (and thus the 8 "zones"; see Figure 1). The objects or things can be anything we can be aware of, including processes, patterns, abstractions, etc.

Below we give a very brief overview of the structural skeleton of the STAGES model. We don't get into describing each level, and skip much about the nuances of the model. More complete descriptions can be found in O'Fallon (2011, 2013; and Fitch et al. 2016) and at www.pacificintegral.com. Our goal here is to give enough bare bones to help readers understand the terms used in this article. Figure 1 shows the levels (stages) and Tiers of the STAGES Model.

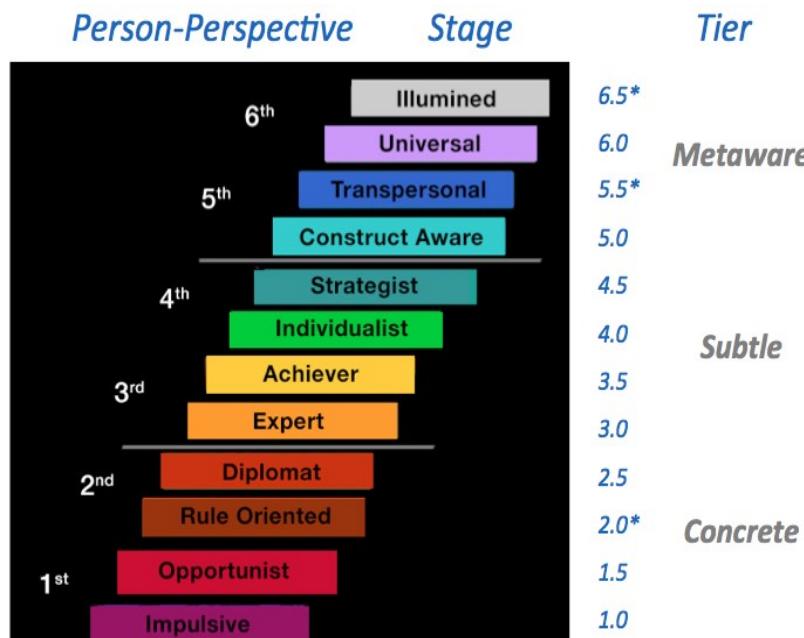


Figure 2: STAGES Levels (stages), Person-Perspectives, and Tiers

The STAGES model posits that the sequence of developmental levels discovered by Loevinger and extended by Cook-Greuter can be explained using the three "primordial" dimensions or ontological polarities of the AQAL model, plus one additional dimension (Tier). It also proposes that the sequence of person-perspectives (PPs), which Cook-Greuter used to structure Loevinger's sequence, can be explained using those four dimensions. The three AQAL dimensions, or primordial polarities, are: Individual/Collective, Inside/Outside, and Interior/Exterior—the familiar concepts defining the AQAL quadrants/zones (Figure 1). The fourth, additional dimension is Tier.

An asterisk in Figure 2 indicates levels STAGES adds to Cook-Greuter's model. It refines three prior categories, splitting them into two: Diplomat becomes 2.0 and 2.5, Construct Aware becomes 5.0 and 5.5, and Unitive becomes 6.0 and 6.5. For most of Loevinger's/Cook-Greuter's levels, the passive-to-active dimension of STAGES fit perfectly, but for these three categories contained both inside (passive) and outside (active) phases (according to O'Fallon's analysis).

In Wilber's model the "Kosmic Address" of any object of inquiry is defined by the values of the three AQAL dimensions. For example, to look at a football team's beliefs objectively are to look at the interior of a collective from a third person (outside) perspective.

When used in STAGES these three dimensions take on a slightly different meaning than their use in AQAL.^[39] In using these constructs to describe the drivers of human development, they must be interpreted a bit differently than their use in AQAL as primordial dimensions applying to all reality. O'Fallon uses the words Passive/Active for Inside/Outside views. These terms refer to the "seeing as" in a subjective way, vs. "looking at" in an objective way, and also correspond roughly with taking a first-person vs. a third-person perspective.^[40]

Meaning-making development is basically about the *types* of objects one is aware of and reasons about, and the *complexity* of those objects and their relationships in thought. The STAGES model posits that, in a general way, meaning making development moves in a particular way within these dimensions of polarities. We start by describing how STAGES uses the primordial dimensions of AQAL to describe development, and then describe the remaining dimension (the Tier: Concrete, Subtle, or MetAware).

Below we give a short introduction to the developmental aspects of these four dimensions, and then go into more depth about each in the next section.

I/C: First, Individuals are understood before Collectives. It is easiest to see that awareness of individual things is simpler than awareness of collectives, groups, or systems of those things. One comprehends persons before families or cultures, cows before herds, planets before solar systems. (There will be exceptions to all of these principles—and we do not have the space in this short overview for a full explanation with all the nuances.)

P/A: Second, Passive (inside, first person) understanding comes before Active (outside, third person). That is, when we first learn about a new type of thing it is an accomplishment to just notice, to be aware of, and be able to talk about, that kind of object. We are in a more receptive relationship to the thing and its appearance happens *to us* in sense. With practice and learning one becomes better at identifying instances of that object and describing its properties and variations. Eventually one takes on a more *active* relationship to the object. One begins to organize and prioritize instances of it, and actively use or refashion ones understanding of it.

E/I: Third, Exterior manifestations of a thing are understood before Interior manifestations. The hypothesis here is that one notices things outside of one before being able to notice them inside. For example, one learns about what it is to be "mad" (concrete tier) or prejudiced (subtle tier) by learning about the characteristics one observes in others; and only later can one turn that understanding inward to ask "am I feeling mad?" or "am I being prejudiced"? The sequencing of this third polar pair is the least important (most refined) in STAGES. It does not factor into differentiating among the levels (when it is used it marks *early* vs. *late* phases within a level). It is also the most hypothetical and it is not addressed by the empirical STAGES research to date.

C/S/M: STAGES adds a fourth (non-AQAL) dimension, called *tiers*, that has three categories: Concrete, Subtle, and Metaware (actually it is the *primary* dimension in STAGES but we introduce it last because we started from the AQAL dimensions). There is a developmental ordering here also. In general we

perceive Concrete objects first, Subtle ones later, and Metaware next. The move from Concrete to Subtle parallel's Piaget's cognitive developmental theory, which shows how development moves from concrete to abstract and formal "operations". Since ego development (meaning-making) is explicitly about I/we/it interrelationships, STAGES (like all construct-developmental models) broadens the Piagetian cognitive scope of development to include inter- and intra-personal themes.

Priorities of the four dimensions.

We have described four dimensions—the tiers have three categories and the other three dimensions are in polar pairs of categories: i.e. Concrete/Subtle/Metaware, Individual/Collective, Passive/Active, and Exterior/Interior. As described above, STAGES posits that development progresses *within* each of these dimensions in the order listed (C before S and M, P before A, etc.). But there is an important relationship *among* these dimensions. It is not that development progresses through all of these dimensions in parallel at the same time, but rather that they have a specific *nested* relationship to each other. For example, there is a movement from Individual to Collective *within each* of the three tiers: Concrete Individual to Concrete Collective, Subtle Individual to Subtle Collective, and Metaware Individual to Metaware Collective. So the C/S/M distinction is the major differentiation of development into three categories, and the I/C is a second level categorization refining this into six categories. The Tier and I/C dimensions define the six person-perspectives, as illustrated in Figure 2 and Table 2.

At the next level of refinement is the Passive to Active movement *within each* of the six person-perspectives, which results in the final 12 developmental levels of the STAGES model. We will ignore the dimension of Exterior/Interior for now, which creates a further refinement of each level into an Early and a Late sub-stage of each level, allowing for a total of 24 sub-stages. This level of refinement is not investigated in our research, is skipped in most STAGES workshops, and is not as often used in SCT scoring (and thus gaining evidence for this aspect of the model is still in process).^[41] Thus, sometimes we will mention the 3 dimensions in the STAGES model (as in Table 2), because that is all that is needed to determine the level and to compare with prior SCT model; and sometimes we will mention the 4 dimensions in the STAGES model, including interior/exterior (as in Figure 3), because the full theory uses four and the full four allows for comparison to the AQAL model.

In positing this clear nested structure STAGES is a simplification of a phenomena, meaning making or perspective taking, that is much more complex. Undoubtedly these four drivers work in a rather messy way that is somewhat nested as in the theory but they also work somewhat in parallel (as well as in ways not addressed by the theory at all). The question of whether the nested sequence is true or valid is what was tested in the most recent empirical research of STAGES (O'Fallon et al. in submission). If a model using the three nested dimensions (drivers) replicates the existing methods of measuring the developmental sequence as defined by Loevinger and Cook-Greuter (which have been validated expensively), then we can assume that this model is a fair approximation of reality (for certain purposes; or at least a fair replication of prior ego development and meaning-making maturity models).

Table 2. STAGES Tiers & Repeating Principles for the 12 Levels

<i>Person Perspective</i>	<i>Name</i>	<i>Tier</i>	<i>I/C</i>	<i>P/A</i>	<i>Complexity</i>
CONCRETE TIER					
1.0	Impulsive	Concrete	Individual	Passive	Receptive
1.5	Egocentric	Concrete	Individual	Active	Active
2.0	Rule oriented	Concrete	Collective	Passive	Reciprocal
2.5	Conformist	Concrete	Collective	Active	Interpenetrative
SUBTLE TIER					
3.0	Expert	Subtle	Individual	Passive	Receptive
3.5	Achiever	Subtle	Individual	Active	Active
4.0	Pluralist	Subtle	Collective	Passive	Reciprocal

4.5	Strategist	Subtle	Collective	Active	Interpenetrative
METAware TIER					
5.0	Construct Aware	MetAware	Individual	Passive	Receptive
5.5	Transpersonal	MetAware	Individual	Active	Active
6.0	Universal	MetAware	Collective	Passive	Reciprocal
6.5	Illumined	MetAware	Collective	Active	Interpenetrative

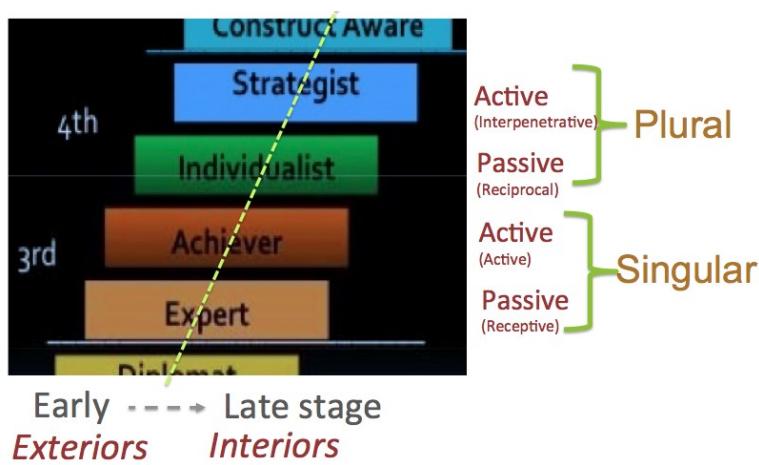


Figure 3: STAGES AQAL-like Structural Dimensions

Further Explanation of the Four Dimensions

Concrete/Subtle/Metaware.

Concrete objects are things one perceives with (or indirectly through) the senses, including not only singular objects such as dogs, noses, planets, atoms, and clouds, but collective things like cities, forests, solar systems, and families. *Subtle* objects of awareness include one's interior thoughts, plans, values, and assumptions; and abstract things like theories, systems, governments, decades, monetary systems, etc.

Note that one can have a concrete understanding of things like "money," "my country," or "living things," or a more abstract, formal, or systems-based understanding of those same things. Even young children understand what money is and how it is used at a concrete level, but more sophisticated thinkers also understand money in terms of the abstract (subtle) concepts of currency, debt, regulations, class, etc.

Also, within the STAGES model, emotions can be understood in concrete *or* subtle ways, where concrete emotions are the basic ones seen in the physical expression of the face or body (mad, sad, scared, happy, etc.), and subtle emotions require an additional depth of inference about what is going on in the interior of another (e.g. jealousy, bliss, disorientation, paranoia, curiosity, etc.). Developmentally, one can imagine a 5 year old child describing how she or another is feeling mad or sad, but probably not jealous or paranoid.

The Metaware tier is more difficult to describe (and is less well understood because there are far fewer individuals at those levels). Metaware awareness includes being aware of awareness itself; having a deep understanding and immediate perception of how one's mind (mind/body) is functioning. This includes an in-the-moment perception and understanding of: how emotion, desire, aversion, and bias can influence thought; the limitations of reason and language; and the machinations of the ego as it attempts to construct the familiar self, and as it tries to construct stable meaning in the face of disorienting

information. One can be aware of projection in the moment as it happens. One becomes comfortable with paradox and unknowing at this tier. Metaware perception and understanding is thought to develop as consciousness increasingly penetrates the processes of thought, feeling, and the subtle sensations moving within the body.^[42]

Individual/Collective.

To understand a collective is to understand both the relationships between objects, and the whole comprised by those related objects. It makes intuitive sense that one usually comprehends individual things before collections. One comprehends persons before families or cultures, cows before herds, planets before solar systems.^[43] Collectives are not simply collections of things grouped together, but (usually) complex systems of interactions in which properties of the group emerge that are distinct from properties of the individual. Molecules have different properties than atoms, people have different properties than cells, and schools of fish have properties beyond those of individual fish.

In the Concrete Tier it is relatively straightforward to understand development as a progression from awareness of individual (1st PP) to collective (2nd PP) objects. The subtle tier is more difficult (subtle!) to explain. At 3rd PP (subtle individual objects) one is aware of ideas, values, plans, etc. To become aware of subtle collectives (4th PP) is to become aware of how groups and cultures work, and how the beliefs and habits of one's culture in part pre-determine an individual's beliefs and habits in a reciprocal relationship (ideas, values, assumptions, etc.). One develops an increased concern with cultural biases and social justice. One sees how "the system" is just as (or more) responsible for people's life circumstances than their individual decisions.^[44] In the later half of the 4th PP one begins to understand human development itself, in both individual and societal manifestations, and one develops a more empathetic understanding of how people's meaning making capacity ("action logic" or perspective taking skill) constrains the complexity of one's world-view.

The progression from individual to collective at the Metaware tier is even more difficult to describe. But briefly, it involves a movement from being identified with the individual self, culture, and species, to awareness becoming one's identity—an identity that holds the collective of living things, or even all of reality.

Passive/Active.

The next dimension of refinement is to subdivide the 6 person-perspectives, defined by the C/S/M and I/C dimensions, into Passive (receptive) and Active sub-levels, resulting in the 12 STAGES levels. With the move to each new person perspective one becomes aware of a new type of object (Concrete Individuals, Concrete Collectives, Subtle Individuals, Subtle Collectives, Metaware Individuals, Metaware Collectives). The first phase of encountering a new type of object is a receptive or passive stage in which it is an accomplishment to just notice, to be aware of, and be able to talk about, that kind of object. With practice and learning one becomes better at identifying variations of this type of object and describing their properties and differences. This phase is called receptive because one is mostly trying to be open to experience and learning—trying to get a better grasp of these new objects.

After sufficient experience and learning in this receptive phase one becomes more *active*—in several ways.^[45] One not only recognizes objects but also begins to organize and prioritize them mentally. In addition, one becomes more active in one's attempts to investigate, discover, and later manipulate and change, and later possibly design or create new objects of this type. Eventually one can become so facile with the skill of working in an area that it becomes unconscious, automatic, and ordinary; and with more practice one can enter "flow states" of efficient mind/body/emotion enacting. (This theme is revisited in the section on "Complexity and Skill Theory" below.)

For example, a young person may learn about the concept of "prejudice." They begin to (receptively) notice or be taught different contexts in which it occurs—they begin to recognize certain behaviors as

prejudiced vs. others that are not. As they mature they begin to appreciate the social and cognitive processes that produce prejudice, develop ideas about what might prevent it, and discern nuances between different types of prejudices and prioritize which are most malevolent. Finally they might engage in programs to ameliorate it or develop their own theory about prejudice. With more practice, one's facility and skill working in this area can lead to one being sought out as an expert who's deep knowledge flows intuitively.

This illustrates the spectrum from passive (or receptive) to increasingly active ways of knowing a thing (though later automated skill and flow states are not *consciously* experienced as "active", at these advanced stages of expertise there is, paradoxically, a return to what feels like a passive engagement or flow). This sequence is not strict of course, as even a very young person can join a campaign to reduce prejudice, though, they may have only a simplistic idea of how it works.

The development from a passive (receptive) to active (engaged) relationship to objects that is seen in the individual ("top") half of each tier repeats at the collective ("bottom") half of each tier. Again, one first efforts just to recognize where, when, and how the new type of collective thing occurs in the world, and eventually understands enough to engage actively with rigorous inquiry, prioritization, manipulation, creation, etc.

But O'Fallon introduces a nuance into this developmental progression as the passive-to-active pattern moves from the individual to the collective phase. One can notice this nuance in our example of prejudice. Prejudice may be understood as a (subtle) property of an *individual*, and indeed one can see that individuals at 3rd PP become concerned with things like prejudice, human rights, fair markets, and democratic decision-making. They tend to see these in terms of benefits and responsibilities of *individuals*. At a more sophisticated level, at a 4th PP, one increasingly understands the *collective*, cultural and social-systems aspects of prejudice (as well as understanding more deeply how not only *others* but *I myself* unavoidably harbor prejudices).

Complexity: R/A/R/I.

O'Fallon describes a 4-part progression within each tier, from passive-individual, to active-individual, to passive-collective, to active-collective, using the terms *Receptive* (passive), *Active*, *Reciprocal*, and *Interpenetrative* (R/A/R/I). This is an alternative vocabulary for talking about the same 4-step move—but the new vocabulary has explanatory power. We could call this the four "complexity" levels within each tier. That is, the passive-collective involves an understanding of *reciprocal* relationship (how objects inter-relate with or co-define others), and the active-collective involves an even deeper *interpenetrative* understanding of how whole systems of objects are intermeshed.

Thus developmental level is identified by asking either three questions: Tier (G/S/C), Individual/Collective (I/C), and Passive/Active (P/A), or an equivalent two questions: Tier (G/S/C) and complexity (R/A/R/I.)

Above we focused on a description of the dimensions of the STAGES model that define its 12 levels. This description does not even begin to describe the rich field of application and insight that the model allows. It is being used to understand leadership development, therapeutic and shadow work, relationship dynamics, and cultural and political trends. All of this can be said also of the ego-development framework, of Kegan's Orders of Consciousness model, and Graves's Spiral Dynamics model. Without critiquing or detracting from the significant insights that these other models contain, the innovations within the STAGES model add important insights to this field of theory and application. For example, the "quadrant" structure that repeats in each of the three tiers supports insights about repeated patterns, for example ways that 1.5, 3.5 and 5.5 show similar dynamics repeated over the span of development (or how shadow issues at 1.5 can assert themselves at 3.5 or 5.5 if they are not integrated).

STAGES vs. Prior Developmental Frameworks

STAGES vs. other SCT assessments.

To summarize, the STAGES model posits that one can use a small set of concepts (C/S/M, I/C, P/A) to describe the developmental drivers underlying the Cook-Greuter/Torbert/Loevinger progression of levels and person-perspectives (levels that are assumed in the field to also map well to Kegan's Orders of Consciousness levels of meaning making maturity). The STAGES model has 12 levels, vs. Cook-Greuter's 9 levels—it subdivides three of Cook-Greuter's levels into two additional levels. These additional levels are defined based on the architecture of the STAGES model (see the asterisks in Figure 1).

As mentioned, in STAGES each level is defined by only three parameters. For example the Achiever level is defined as Subtle-Individual-Active (see Table 2). Trained STAGES scorers base their scoring of each SCT stem completion on these concepts.^[46] This contrasts sharply with how completions are scored in the Loevinger tradition, which uses an example-based method. That is, scoring generally consists of matching the completion within a very large set of exemplars categorized according to each level (and also sub-categorized for each stem, and further categorized thematically). Cook-Greuter's and Torbert's scoring systems, following closely from Loevinger, is exemplar-based, and allows for a small amount of consideration of the generic structure or nature of the completions.^[47] STAGES scoring is not exemplar-based and looks almost exclusively at the underlying structure and nature of completions using the general principles described above.

One of the innovations of STAGES over prior ego-development frameworks is that it purports a systematic underlying structure to explain the development of meaning making. STAGES has been metaphorically compared to the periodic table of the elements. Prior to the Periodic Table, scientists understood the elements in terms of their surface properties, such as density, color, conductivity, magnetism, melting point, reactivity, etc. They could be arranged and ordered and compared in terms of these properties but the understanding of each element was focused descriptively on its own unique set of properties. The modern understanding of atoms as being composed of electrons orbiting a nucleus allowed scientists to explain the properties of elements in terms of underlying causal mechanisms and structures, leading to the modern organization illustrated by the Periodic Table. Chemistry moved from being primarily a (data-driven) descriptive process to a theory-driven one based on underlying principles from particle physics. One of the stark achievements of this advance was that scientists could predict that certain elements must exist even though they had never been observed, based on locations in the table where nothing had been observed yet.

STAGES is *metaphorically* similar in that it moves the assessment of ego development from a descriptive method based on the properties of each level (and the specific examples for each level given in the scoring manual) to a basis on underlying structure.^[48] The structure of the Periodic Table is based on the two categories of number of neutrons and number of elections, whereas the structure of STAGES is based on the four categories of C/S/M, I/C, P/A, and Int/Ext.

Two things are important to note in comparing STAGES with other (data-driven) variations on Loevinger's work. First, Loevinger's and Cook-Greuter's model and the scoring manual *do* mention some underlying principles and mechanisms thought to drive development, so they are not *exclusively* tied to surface features or to matching with exemplars. But in general these underlying principles are held lightly and not used to propose a formal underlying structure.^[49] Second, O'Fallon's descriptions of each stage are founded upon her three main structural dimensions, but both the theory and the scoring manual do not limit the description of each level to these dimensions, and they also mention some surface properties.

A key difference is that STAGES purports to be domain-independent. Its underlying developmental principles can be applied, theoretically, to any area of meaning-making. The scoring system is based on underlying language structures and markers, rather than on (categorized sets of) exemplars. This allows STAGES applications to break away from the quasi-standard set of 36 sentence completions and to be applied to specific areas such as leadership, relationship, collaboration, money and business, ethical reasoning, education, etc. Whereas Loevinger's work was specifically about "ego development" defined as a "master trait" applying to one's relationship to reality in general, the STAGES assessment can (theoretically) be used to assess meaning-making in specific domains. This is done by tailoring the sentence stems to each domain.^[50] To the extent that more new sentences are introduced, it becomes

increasingly difficult to statistically compare a STAGES assessment to ego development assessments scored within other SCT models (though current STAGES "specialty protocols," which have only 6 modified stems, have proven to correlate very well with prior stems).

An additional benefit is that STAGES is not only unconstrained by the particular set of sentence stems, but it can move outside the sentence-completion method entirely and applied to any text, including essay responses, blog posts, books, lectures, etc. O'Fallon has recently begun to experiment with such applications. Here, again, as these other sources of text are not (usually) "projective" tests, one must be cautious in assuming that research based on the SCT is transferable to the next context, and new research will be needed to test validity.

As a model STAGES is more elegant, parsimonious, and deep (it proposes underlying causal mechanisms) than the models it was derived from. Sounds good, but is the STAGES model *valid*? In fact it may sound too good, too simple and elegant, or too simplistic and abstracted, to be true—in boiling the complex field of meaning-making development down into three or four primary dimensions. Only empirical research can answer this question. The confirming results of O'Fallon's first empirical study are summarized later.

Theory Comparison Caveats.

In recasting the familiar developmental levels of Loevinger and Cook-Greuter into the highly structured STAGES model, there are epistemological tradeoffs. We must be careful, in this and all models, not to confuse the map with the territory. The territory, human cognition, meaning-making, and mind/body /spirit presence, like most of nature, is highly complex, varied, and resists being reduced to simple categories. All (valid) models are true but partial. The STAGES model is simpler, has less major moving parts, so-to-speak, than prior theories. To the extent that it is accurate it can add clarity and power to our understanding, as the periodic table did with chemistry. But the human condition is vastly more complex than the structure of the atom, and the categories of any model that tries to capture its intricate nature will fail to some extent. Those using STAGES or any model of human nature should be vigilant in questioning what gets left on the cutting room floor when the clean categorical slices of the theoretical model are used to interpret reality (I give similar cautions about the AQAL model in Murray, 2011). O'Fallon supports this caution.

In any model-using community, that which does not easily fit the model is difficult to see and it can become marginalized. For example, after IQ assessment was standardized, many types and nuances of intelligence may have gone unnoticed in the excitement of using such a powerful tool. Society came to focus on, valorize, and prioritize the types of intelligence that could be measured—the tail wagging the dog in a way that is, unfortunately, not uncommon, because any tool that reduces the (real) complexities of human life is a welcomed (and dangerous) relief.

STAGES proposes a set of three (or four) dimensions as the main drivers and underlying structures of meaning-making development. Strong concordance (or "replicability," described later) with Cook-Greuter's (and thus Loevinger's) model indicates that there is considerable validity in this hypothesis. O'Fallon shows evidence that its measurement is comparable to prior models. We have argued that STAGES is measuring something very similar to, but not identical to, prior models. O'Fallon, in our reading, is *not* claiming that human meaning making can be fully explained or captured within the STAGES model, but that enough is captured to support the model's usefulness.

So, in looking at the human condition through this new lens, we can ask: What properties are shaved off of the earlier model's descriptions for each level, and what properties are added? More research is needed to answer that question. And as we learn more to answer that question, then we must ask about the practical consequences—the benefits and tradeoffs (the liberations and calamities!)—of the, perhaps slight, pivot in the meaning of meaning-making that the new model imposes. Just as Kegan's model points to something a bit different than Cook-Greuter's, O'Fallon's model ("map") points to territory similar to, but different than, Cook-Greuter's.

STAGES, States, and cognitive/brain theories

STAGES lends itself better than its predecessor theories to coordination with contemporary cognitive and brain theories. The C/S/M sequence is in line with neo-Piagetian cognitive and construct-developmental theories. In the next section we show how STAGES has striking similarities with Hierarchical Complexity Theory and Skill Theory, the most recent neo-Piagetian models. The progression of "complexity" within each tier (combining the I/C and P/A dimensions) also parallels contemporary understandings of development.

To discuss another connection between O'Fallon's theory and cognitive psychology we need to bring in the notion of State, which, in integral theory, is contrasted with Stage. In Wilber's model, the "states" are drawn from "the great wisdom traditions (such as Christian mysticism, Vedanta Hinduism, Vajrayana Buddhism, and Jewish Kabbalah)" which "maintain that the 3 natural states of consciousness—waking, dreaming, and deep formless sleep...contain an entire spectrum of spiritual enlightenment" (especially if a fourth non-dual state is added) (Wilber, 2006, p. 4).

The "Wilber/Combs Matrix" principle claims that any of the states of consciousness can be experienced at any of the stages of development, but that the *interpretation* of the experience of a state comes through one's stage of development (Wilber, 2005). O'Fallon agrees, but also believes that experiencing certain states of mind are rough (perhaps not strict), prerequisites for the achievement of certain stable stages of development.^[51]

Here I offer a definition of states and stages that is compatible with, but different from, Wilber's description of the Wilber-Combs Matrix, that allows the STAGES model to coordinate with contemporary cognitive and brain science. "State" points to any experience one is having in any moment, and "Stage" points to stable developmentally attained patterns of behavior, understanding, or knowledge. To put it succinctly: states are about how brains *fire*, and stages are about how brains are *wired*. Granted, this model is within a rather reductive brain-centric view of consciousness, but the fire/wire metaphor can be easily extended to less reductive whole-body or meta-body theories of consciousness and serve the same function.

One's state is defined by which elements (nodes, neurons, ideas, etc.) are currently activated (and the firing activations and de-activations between nodes in any moment). A Stage, as wiring, refers to pathways of *potential* activation.^[52] Though neurons generally fire in accordance with their most robust wiring connections, extreme conditions, mind-altering drugs, etc. can create firing patterns that are atypical, perhaps causing signals to travel down weak wiring links to create novel firing patterns, non-ordinary experiences, states, or new ideas.^[53]

Non-ordinary states become ordinary if repeated over time. The "wow," "aha," or spiritually heightened experience of some states may be more attributable to their *novelty* than their content. These experiences may actually be a pleasant type of cognitive disorientation, and the brain seems wired to give a boost of pleasure rewarding novelty, openness, and exploration (Panksepp, 2005). Spiritual adepts tell us that the specialness of new states, stages, or insights becomes ordinary or normalized over time. The experience of an infant first discovering that *that thing* moving in my field of vision is *my hand*, or getting the first flashes of object permanence; or as a child, discovering that the letters of the alphabet can be put together to actually *mean* something; or, later in life, realizing that the self is an illusion—all of these experiences may create a very similar state of expansive arousal that diminishes over time and practice, as the learning becomes a stable wiring pattern.

The punch line to this line of reasoning is the common understanding in cognitive neuroscience that "*nodes that fire together wire together*" (i.e. associations are strengthened through repeated use) and "*nodes that are wired together tend to fire together*" (in cascades of association and patterns of reinforcement and inhibition) (Sompolinsky, 1987; Siegel, 2012). This model is compatible with Wilber's notion that states can be experienced but stages can't—that they must be inferred from studies of behavior. Stages point to modes of potential activation leading to experience and behavior—they can not be directly experienced any more than we can inspect our own mind and see the "rules" of grammar that we use when we speak.

This leads directly to O'Fallon's conjecture that repeated experience of certain states may be necessary (but not sufficient) for attaining certain stages. It is a straight following from common cognitive neuroscience theory, given our more expanded and non-spiritualized meaning of "state."

STAGES vs. Complexity and Skill Theory

At this point it is worth taking a tangent to mention how the STAGES model parallels what has been discovered in the leading edge of "Neo-Piagetian" developmental research. [This sub-section can be skipped if you are not interested in how STAGES compares to developmental theories outside the Loevinger tradition.]

The top researchers in this field are Kurt Fischer, who developed Skill Theory (Fischer et al., 1980, 2008), and Michael Commons, who developed Hierarchical Complexity Theory (Commons et al., 1998).^[54] These theories have made key advances in recent decades by working out domain-independent models of human development. That is, they can theoretically be applied to any skill or capacity, and have been tested empirically in domains including reflective reasoning, mathematics knowledge, physical/athletic skill, emotional intelligence, ethical reasoning, and specific themes of reasoning such as "What is a good life?" Both are theory-driven (as opposed to the data-driven orientation of Loevinger's model), and have been strongly empirically validated. Both include modern interdisciplinary principles from systems theories and brain science (which has not been applied much yet to ego development theories).

Both describe development in terms of an invariant sequence of levels, each of which coordinates skills or capacities of prior levels. Both describe a sequence of about 12 levels that begins with the most basic sensory-motor skills of the infant, and ends with the most complex reasoning seen in adults. Significantly, these two models were developed without knowledge of the other, and later were seen to describe almost exactly the same levels using very similar principles. We will focus on Fischer's Skill Theory.

That the STAGES model has some strong similarities with these models, even though O'Fallon did not intentionally build upon them, is important evidence for the face validity of STAGES. In addition, it positions STAGES as a bridge-building model between research on ego development and meaning-making, and the important work of the Neo-Piagetians.^[55]

The first similarity is that STAGES, like Neo-Piagetian models, uses a theory-motivated measurement and looks to underlying mechanisms that explain the surface features of development—which Loevinger's assessment, being data-driven, does not do directly.^[56] Second, and following from the first, STAGES, like Skill Theory, is relatively domain-independent. We say relatively because, while Skill Theory can be applied to *any* human skill or capacity, STAGES seems to be limited to assessing *meaning making*, but it can be applied to assess meaning making in any number of domains, beyond the single domain of ego development measured by the WUSCT/MAP/GLP SCTs.

Third is the most striking similarity. Both theories describes development in terms a progression of tiers, where each tier refers to a new type of emergent object; and within each tier is a progression of increasing complexity and sophistication in working with that type of object.

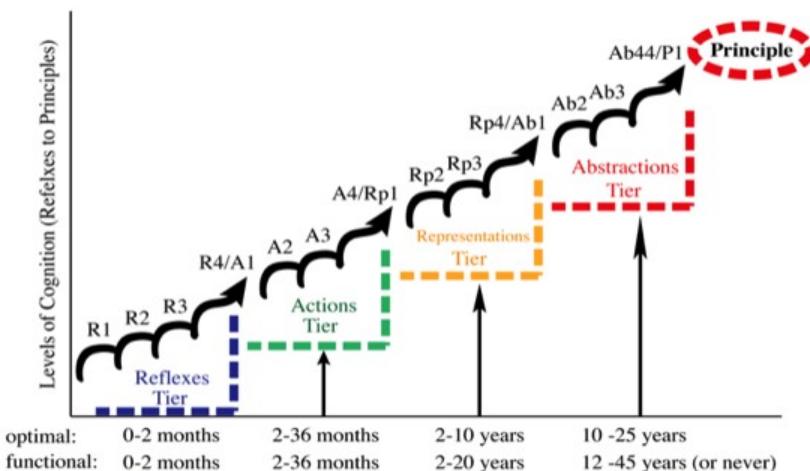


Fig. 1. The 13 levels and 4 tiers of development in Fischer's skill theory

**Figure 4: The 13 levels and 4 tiers of development in Fischer's Skill Theory
(From Murphy, 2015)**

Skill Theory includes 4 tiers (which can be extended to 5) with 3 levels each; while STAGES includes 3 tiers (which O'Fallon sometimes extends to 4, adding a "Unitive" tier) with 4 levels each. The tiers in Skill Theory are Reflexes, Actions, Representations, Abstractions (with the top being the beginning of a Principles tier)—compared with Concrete, Subtle, and MetAware tiers in STAGES. [Appendix 2](#) includes a coordination of the levels of the two theories.

We can see how, in a sense, Skill Theory and STAGES divide up the same conceptual space, describing increasing complexity, into three, vs. four, slices, respectively. Skill theory (and HCC) look specifically at task complexity or idea-complexity, while STAGES also includes an aspect of one's performative relationship to an idea, i.e. whether one has a passive or active relationship with it. The three levels within each tier in Skill Theory are *single things*, *mappings* of things, and *systems* of things. To understand single things is very similar to what O'Fallon calls the passive (receptive) phase of understanding a new object—in which one is challenged to just identify single instances of the thing. At the mapping phase of Skill Theory one begins to understand relationships between single objects. One moves from simple relationships between individual objects to more complex multi-object relationships. This overlaps with "active" and "reciprocal" STAGES (while "reciprocal" points to more than just relationship, i.e. to *inter*-relationship). At the *systems* phase one perceives entire systems of relationships, where objects may inter-influence and co-define each other, which again, is close to the interpenetrative in STAGES (while "interpenetration" points to something a bit more integrated than a system of relationships).

Skill theory has it that the first level of any tier is also a *fourth* level of the prior tier (see Figure 4). The fourth level is where the systemic understanding becomes so dense that a new single object (or type of object) emerges at the next tier. This is even closer to O'Fallon's use of an "interpenetration" of objects.

The two models are not expected to line up exactly, as they are describing somewhat different phenomena. Skill Theory is a theory of any skill, which has mainly been applied predominantly to cognitive domains. It is best used to measure very specific (narrow) lines of development, in which the way each specific sub-skill builds upon prior skills can be mapped out precisely through analysis of a task. Skill Theory can be used to target very specific skill progressions, as in the understanding of physics curricula, strategic thinking in organizations, ideas about "the good life," or a cognitive understanding of the concept of self (e.g. Dawson, 2004).

STAGES is a theory of meaning-making, which explicitly includes self/other/world relationships as applied to *any* life-theme, and which thus covers a rather *broad* and inclusive "line" of development (as is Kegan's theory). In being holistic, meaning-making assessment is expected to be a bit fuzzier and imprecise, compared with applications of Skill Theory. Skill Theory measures the progression of *conceptual* understanding and *logical* complexity. STAGES assessment looks at logical complexity, but also includes interpersonal/relational factors and the being/doing elements of passive vs. active

relationship to the object.^[57]

Our purpose here is not to compare the models level-by level, or to compare the results of each research lineage, but to note the similarities that (a) support the face validity of STAGES, and (b) point toward future analysis and synthesis between the models (a project left for another day). One can make rough correspondences between the levels of each, but in doing so the boundaries do not line up precisely. The decision to break up the sequence within each tier into three vs. four levels is strongly grounded in the underlying theory of each model, yet can also be described as the somewhat arbitrary attempt to "cut the pie" of a complex phenomena using different conceptual slices. Both try to use a consistent "measuring stick" to define a standardized transition between each level (in contrast, the Loevinger model does not have a mechanism for trying to argue that jumps between each level are of comparable size).

Validity Studies of the STAGES Model

It is one thing to have an intuition, as O'Fallon did, that the complex set of properties defined for Loevinger's and Cook-Greuter's levels of ego development might be captured with three or four dimensions—using the very dimensions that form Wilber's AQAL model of ontological dimensions of reality. This intuition was based on a significant amount of anecdotal data, i.e. O'Fallon's experience scoring within Cook-Greuter's model over many years. But it may seem quite a stretch to hypothesize that the new system (the STAGES model and scoring method) measures the same thing as the prior systems (the ego development model and the MAP scoring method). STAGES has proven itself to be very useful at the level of metaphor and model, i.e. as a meaning-generative tool in workshops, coaching, personal growth, and psychotherapeutic understanding. It might be an elegant theory that sounds good but throws out too much in the transition to an elegant structured model. In fact, empirical research has shown that STAGES is a valid measure of ego development (meaning making maturity), comparable with prior assessments.

Below I summarize the research reported in (O'Fallon, in preparation), without giving too much detail due to concerns about the need for peer reviewed scholarly articles to be original and not available prior in the public domain. A shorter version of this summary also appeared in O'Fallon (2013). (Note, the remainder of this section duplicates text from the Response article in this ILR issue.)

O'Fallon and colleagues conducted an empirical study of the validity of STAGES using a total of about 150 inventories (of 36 sentence completions each). Two major hypotheses were tested. The primary question was whether scoring with the STAGES method replicated scoring by prior Cook-Greuter/Loevinger methods ("CG/L"). Success of replication would support the theoretical model underlying the STAGES scoring system, which posits that the three dimensions of Concrete/Subtle/MetAware, Individual/Collective, and Passive/Active, can be used to construct an underlying model of the perspectival drivers of meaning-making (ego-) development.

This primary hypothesis was tested for levels up to Strategist (" $<=4.5$ "). Starting with the MetAware level (5.0, 5.5, 6.0, 6.5), STAGES has four additional levels corresponding to Cook-Greuter's two stages (Construct Aware and Unitive). This, and the paucity of data on the MAP side, makes the systems difficult to compare using replicability methods. Also, the definitions of the levels diverge more between the systems at these levels. For these highest levels an inter-rater-reliability study was done. Thus the second main hypothesis was that STAGES scoring at these levels was valid by having adequate inter rater reliability (IRR). Proving the first hypothesis demonstrates that the repeating pattern (P/A/R/I) is valid for the first two tiers, and this suggests that the pattern will repeat itself in the third tier.

Each inventory had been previously scored by one of a set of four certified scorers using the CG/L method (some random inter-rater checking was done with this sample). Each inventory was then scored by three trained STAGES scorers (assigned from a group of four trained scorers). The use of four scorers represents a higher level of rigor than is found in most replication or IRR studies.

Samples were drawn from over 1000 inventories previously scored using the CG/L method (O'Fallon, 2013). Based on statistical power analysis on the required sample size for the desired confidence level, about 150 protocols were randomly sampled. These included about 75 inventories for the replicability

study of the <=4.5 levels, and about the same number of inventories for the IRR study of the >=5.0 levels. For the >=5.0 inventories all available data was used. For the <=4.5 inventories a stratified random sampling method was used to insure: (1) a relatively equal number of protocols at each level, and (2) an even representation across CG/L scorers (as much as possible).^[58]

The match (replicability) between STAGES and CG/L (<=4.5) was assessed in multiple ways. Because O'Fallon was a scorer on both the CG/L side and STAGES side (having scored some of the inventories in the data set before the STAGES model was created) and because she is the creator of the STAGES model, special precautions were taken to ensure validity. First, an expert inter-rater was used to randomly cross-check O'Fallon's CG/L-system scores. Second, each of the four scorers was compared separately with CG/L scores.^[59]

The replicability of all four STAGES scorers (for levels <=4.5) was found to be in the "excellent" range.^[60] Thus confirming the primary hypothesis of the study.

For the IRR study of the >-5.0 levels, agreement was in the "substantial" range, which is quite good considering that stages tend to get more difficult to score at later levels. Thus the second hypothesis was confirmed. We can also note that the IRR for the entire range of levels take together is in the "excellent" range over the four raters. Many additional details of the background, method, and statistical analysis are described in O'Fallon (in preparation), and this summary is *not* meant to be a sufficient argument for the validity of the STAGES model or scoring system. This summary merely points to a to-be-published paper that argues for the validity of the model and scoring system.

Conclusions

I hope this paper has been useful to those wanting a deeper understanding of wisdom skills, adult developmental assessment and sentence completion tests, the STAGES model, and/or the scientific validity studies supporting them.

Research on STAGES is in early phases. The research to date seems quite strong, and we believe strong enough to support its use in many contexts—but many more studies are possible and needed to establish the theory at the level of robustness of its predecessors. Also, it never hurts to revisit the list of "Caveats and Concerns" at the end of the Introduction Section. The business of assessing and rating someone's level or stage of development is approximate and prone to pigeonholing, misinterpretation, and even misuse. Such assessments powerful research models for studying trends in populations, but for use with individuals the suggested use is as a starting point for interactive dialogue and self-reflection that does not take the assessment as the final word.

References

- Al-Owidha, A., Green, K. E., & Kroger, J. (2009). On the question of an identity status category order: Rasch model step and scale statistics used to identify category order. *International Journal of Behavioral Development*, 33(1), 88-96.
- Alexander, C. N., Heaton, D. P., & Chandler, H. M. (1994). Advanced human development in the vedic psychology of Maharishi Mahesh Yogi: Theory and research. In M. E. Miller & S. Cook-Greuter (Eds.), *Transcendence and mature thought in adulthood* (pp. 39-70). New York: Rowman & Littlefield.
- American Association of Colleges and Universities (AACU, 2007). College learning for the new global century. Washington, DC: AACU.
- Anderson, J. (1983). The architecture of cognition. Cambridge, MA: Harvard Univ. Press.
- Baltes, P. B., & Staudinger, U. M. (2000). Wisdom: A metaheuristic (pragmatic) to orchestrate mind and virtue toward excellence. *American Psychologist*, 55(1), 122.
- Bangen, K. J., Meeks, T. W., & , D. V. (2013). Defining and assessing wisdom: A review of the literature. *The American Journal of Geriatric Psychiatry*, 21(12), 1254-1266.
- Barker, E. H., & Torbert, W. R. (2011). Generating and Measuring Practical Differences in Leadership Performance at Postconventional Action-Logics. *The Postconventional Personality: Assessing, Researching, and Theorizing Higher Development*, 39.
- Beck, D. E. & Cowan, C. C. (1996). *Spiral Dynamics*. Malden, MA: Blackwell Publishers, Ltd.

- Binder, T., & Kay, J. (2011). Ego development: A key aspect of personality development. Leadership and personal development. A toolbox for the 21st century professional, 52-53.
- Black, L., Welser, H., Cosley, D., and DeGroot, J., Self (2011). Governance Through Group Discussion in Wikipedia Measuring Deliberation in Online Groups. *Small Group Research* 42(5) pp. 595-634.
- Blumentritt, T. (2011). Is higher better? a review and analysis of the correlates of post-conventional ego development. In *Pfaffenberger, A. Marko, P., & Combs, A. (Eds.), The Post-conventional Personality: Assessing, Researching, and Theorizing Higher Development*, 153-162.
- Browning, D. L. (1987). Ego development, authoritarianism, and social status: An investigation of the incremental validity of Loevinger's Sentence Completion Test (Short Form). *Journal of Personality and Social Psychology*, 53(1), 113-118.
- Browning, D.L. (1986). Psychiatric ward behavior and length of stay in adolescent and young adult inpatients: A developmental approach to prediction. *Journal of Consulting and Clinical Psychology*, 54, 227-230
- Bursik, K. (1991). Adaptation to divorce and ego development in adult women. *Journal of Personality and Social Psychology*, 60, 300-306.
- Bushe, G. R., & Gibbs, B. W. (1990). Predicting organization development consulting competence from the Myers-Briggs type indicator and stage of ego development. *Journal of Applied Behavioral Science*, 47, 337- 357.
- Carlson, V.K., & Westenberg, P.M. (1998). Cross-cultural research with the WUSCT. In J. Loevinger (Ed.), *Technical foundations for measuring ego development* (pp. 57-75). Mahwah, NJ: Erlbaum.
- Clark, D. V.D. Sampson, K Stegmann, M Marttunen, I Kollar, J Janssen, A Weinberger, M Menekse, G Erkens, L Laurinen (2009). Scaffolding scientific argumentation between multiple students in online learning environments to support the development of 21st century skills. In *proceeding of: National Research Council Workshop Exploring the Intersection of Science Education and the Development of 21st Century Skills*, At Washington D.C.
- Cohn, L. D., & Westenberg, P. M. (2004). Intelligence and Maturity: Meta-Analytic Evidence for the Incremental and Discriminant Validity of Loevinger's Measure of Ego Development. *Journal of Personality and Social Psychology*, 86(5), 760-772.
- Cohn, L.D. (1998). Age trends in personality development: A quantitative review. In P.M. Westenberg, A. Blasi, & L.D. Cohn (Eds.), *Personality development: Theoretical, empirical, and clinical investigations of Loevinger's conception of ego development* (pp. 133-143). Mahwah, NJ: Erlbaum.
- Commons, M. L., Trudeau, E. J., Stein, S. A., Richards, F. A., & Krause, S. R. (1998). Hierarchical complexity of tasks shows the existence of developmental stages. *Developmental Review*, 18(3), 237-278.
- Cook-Greuter, S. R. (1999). *Postautonomous ego development: A study of its nature and measurement* (Doctoral dissertation, Harvard Graduate School of Education).
- Cook-Greuter, S. R. (2002). A detailed description of the development of nine action logics adapted from ego development theory for the leadership development framework. *Retrieved Oct, 13, 2002*.
- Cook-Greuter, S. R. (2004). Making the case for a developmental perspective. *Industrial and Commercial Training*, 36(7), 275-281.
- Cook-Greuter, S. R. (2011). A report from the scoring trenches. *The Postconventional Personality. Assessing, researching, and theorizing higher development*, 57-71.
- Dawson-Tunik, T. L., Commons, M., Wilson, M., & Fischer, K. W. (2005). The shape of development. *European Journal of Developmental Psychology*, 2(2), 163-195.
- Dawson, T. (2004). Assessing intellectual development: Three approaches, one sequence. *Journal of adult development*, 11(2), 71-85.
- Diamond, J. M. (1998). *Guns, germs and steel: a short history of everybody for the last 13,000 years*. Random House.
- Esbjörn-Hargens, S. (2009). An Overview of Integral Theory: An All-Inclusive Framework for the 21st Century. Integral Institute, Resource Paper No. 1, March 2009, pp. 1-24.
- Fischer, K. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological Review*, 87(6), 477-531.
- Fischer, K. W. (2008). Dynamic cycles of cognitive and brain development: Measuring growth in mind, brain, and education. *The educated brain: Essays in neuroeducation*, 127-150.
- Fischer, K. W., & Farrar, M. J. (1987). Generalizations about generalization: How a theory of skill development explains both generality and specificity. *International Journal of Psychology*, 22(5-6), 643-677.
- Fischer, K. W., Hand, H. H., & Russell, S. (1984). The development of abstractions in adolescence and adulthood. *Beyond formal operations*, 1, 43-73.
- Fischer, K.W. & Pruyne, E. (2002). Reflective thinking in adulthood: Emergence, development, and variation. In *Handbook of Adult Development*, J. Demick & C. Andreeoletti (Eds). New York: Plenum Press.
- Fitch, G. (2016). In, As, and Towards the Kosmic We. Chapter in O. Gunnlaugson & M.Brabant (Eds.) *Cohering the Integral We Space: Engaging Collective Emergence, Wisdom, and Healing in Groups*, p. 79-93. Integral Publishing House.
- Fry, L. W. J., & Wigglesworth, C. G. (2013). Toward a theory of spiritual intelligence and spiritual leader development. *International journal on spirituality and organization leadership*, 47, 1-41.
- Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.
- Gould, S. J., & Eldredge, N. (1977). Punctuated equilibria: the tempo and mode of evolution reconsidered. *Paleobiology*, 3(02), 115-151.
- Harari, Y. N., & Perkins, D. (2014). *Sapiens: A brief history of humankind* (p. 443). London: Harvill Secker.
- Holt, R. R. (1980). Loevinger's measure of ego development: Reliability and national norms for male and female short forms. *Journal of Personality and Social Psychology*, 39(5), 909.
- Hoppe, C. F., & Loevinger, J. (1977). Ego development and conformity: A construct validity study of the Washington University Sentence Completion Test. *Journal of Personality Assessment*, 41(5), 497-504.
- Hoppe, C. F., & Loevinger, J. (1977). Ego development and conformity: A construct validity study of the Washington University Sentence

- Completion Test. *Journal of Personality Assessment*, 41(5), 497-504.
- Hy, L. X., & Loevinger, J. (2014). *Measuring ego development*. Psychology Press.
- Kegan, R. (1982). *The Evolving Self*. Harvard University Press.
- Kegan, R. (1994). *In over our heads: The mental demands of modern life*. Cambridge, MA: Harvard University Press.
- Kegan, R., Lahey, L., & Souvaine, E. (1998). From taxonomy to ontogeny: Thoughts on Loevinger's theory in relation to subject-object psychology.
- King, L. A. (2011). The challenge of ego development. In Pfaffenberger, A. Marko, P., & Combs, A. (Eds.). *The Post-conventional Personality: Assessing, Researching, and Theorizing Higher Development*, 163-174.
- King, P. M., & Kitchener, K. S. (2004). Reflective judgment: Theory and research on the development of epistemic assumptions through adulthood. *Educational psychologist*, 39(1), 5-18.
- King, P.M. and Kitchener, K.S. (1994). *Developing reflective judgment: Understanding and promoting intellectual growth and critical thinking in adolescents and adults*. San Francisco: Jossey-Bass.
- Knowles, M. S. (1970). The modern practice of adult education (Vol. 41). New York: New York Association Press.
- Koestler, A. (1967). *The Ghost in the Machine*. NY: Macmillan.
- Kuhn, D. (1999). A Developmental Model of Critical Thinking. *Educational Researcher*, March, 1999, pp. 16-25.
- Kuhn, D. (2000). Metacognitive Development. *Current Directions in Psychological Science*, 9(5), pp. 178-181.
- Kuhn, D., & Pease, M. (2008). What needs to develop in the development of inquiry skills? *Cognition and Instruction*, 26(4), 512-559.
- Lakatos, I. (1976). *Proofs and refutations: The logic of mathematical discovery*. J. Worrall & E. Zahar, (Eds.). Cambridge, MA: Cambridge Univ. Press.
- Laird, J., Newell, A., & Rosenbloom, P. (1987). "SOAR: An Architecture for General Intelligence," *Artificial Intelligence*, Vol. 33.
- Lambert, H.V. (1972). A comparison of Jane Loevinger's theory of ego development and Lawrence Kohlberg 's theory of moral development. Unpublished doctoral dissertation, University of Chicago, Chicago, IL.
- Lee, P. M. (2012). *Bayesian statistics: an introduction*. John Wiley & Sons.
- Loevinger, J. (1979). Construct validity of the sentence completion test of ego development. *Applied Psychological Measurement*, 3(3), 281-311.
- Loevinger, J. (1985). Revision of the sentence completion test for ego development. *Journal of personality and social psychology*, 48(2), 420.
- Loevinger, J. (Ed.). (1998). Technical foundations for measuring ego development: The Washington University sentence completion test. Psychology Press.
- Loevinger, J., & Wessler, R. (1970). *Measuring ego development: Construction and use o fa Sentence Completion Test*, Vol. 1. San Francisco: Jossey-Bass.
- Lucas, R.H. (1971). Validation of a test of ego development by means of a standardized interview (Doctoral dissertation, Washington University, St. Louis). Dissertation Abstracts International, 32, 2204B.
- Manners, J., & Durkin, K. (2001). A critical review of the validity of ego development theory and its measurement. *Journal of personality assessment*, 77(3), 541-567.
- McCrae, R.R., & Costa, P.T. (1980). Openness to experience and ego level in Loevinger's sentence completion test: Dispositional contributions in developmental models of personality. *Journal of Personality and Social Psychology*, 39, 1179-1190.
- Meeks, T. W., & Jeste, D. V. (2009). Neurobiology of wisdom: A literature overview. *Archives of general psychiatry*, 66(4), 355-365.
- Meizrow, J. (1991). Transformative Dimensions of Adult Learning [Kindle for Mac version]. Retrieved from <http://amzn.to/oxPteI>
- Meizrow, J. (2000). Learning to think like an adult: Core concepts of transformational theory. In J. Meizrow & Associates (Eds.), *Fostering critical reflection in adulthood*. San Francisco: Jossey-Bass.
- Mervis, B. & Rosch, E. (1981). Categories of natural objects. *Annual Review of Psychology*, 32. p. 89-115.
- Muhlberger, P., & Weber, L. M. (2006). Lessons from the Virtual Agora Project: The effects of agency, identity, information, and deliberation on political knowledge. *Journal of Public Deliberation*, 2(1), 6.
- Murphy, R.S. (2015). What Is Fischer's Skill Theory? On the *Neuroscience for Teaching EFL* web site. Figure downloaded 6/1/2015from <http://fab-efl.com/onlinelearning/page21/page19/index.html>.
- Murray, T. (2009). What is the Integral in Integral Education? From progressive pedagogy to integral pedagogy *Integral Review*, Vol. 5 No. 1, June 2009, pp. 96-133.
- Murray, T. (2009b). Intuiting the Cognitive Line in Developmental Assessment: Do Heart and Ego Develop Through Hierarchical Integration? *Integral Review*, Vol. 5 No. 2, December 2009, p. 343-354.
- Murray, T. (2011; and to appear). "Integralist mental models of adult development: Provisos from a Users Guide". In *Integral Leadership Review* Vol. 11, No. 2 March 2011, and to appear as a chapter in Esbjörn-Hargens (Ed) book *True But Partial: Essential Critiques of Integral Theory*.
- Murray, T. (2011). "Toward post-metaphysical enactments: On epistemic drives, negative capability, and indeterminacy analysis" *Integral Review*, Vol. 7 No. 2, October 2011, p. 92-125.
- Newman, D.L., Tellegen, A., & Bouchard, T.J. (1998). Individual differences in adult ego development: Sources of influence in twins reared apart. *Journal of Personality and Social Psychology*, 74, 985-995.
- Nicolaides, A. I. (2008). *Learning their way through ambiguity: Explorations of how nine developmentally mature adults make sense of ambiguity* (Doctoral dissertation, Teachers College, Columbia University).
- Novy, D. M., & Francis, D. J. (1992). Psychometric properties of the Washington University Sentence Completion Test. *Educational and psychological measurement*, 52(4), 1029-1039.
- NSTA (2011). Quality Science Education and 21st Century Skills . Publication of the National Science Teachers Association, 21 Feb. 2011.

- Downloaded 1/10/14/ from http://science.nsta.org/nstexpress/PositionStatementDraft_21stCenturySkills.pdf.
- O'Fallon, T. & Polister, N., Blazej Neradiek, M. (in preparation). STAGES: A New Integral Scoring Methodology for Perspectival Levels in Ego Development.
- O'Fallon, T., Murray, T., Fitch, G., Kesler, J. & Barta, K. (this issue of ILR). A Response to Critiques of the STAGES Developmental Model. *Integral Leadership Review*, Fall, 2016.
- O'Fallon, T. (June 2010a). Developmental experiments in individual and collective movement to second tier. *Journal of Integral Theory and Practice*, 5(2), 149-160.
- O'Fallon, T. J. (2010b). The collapse of the Wilber-Combs Matrix: The interpenetration of the state and structure stages. Presented at the July 2010 *Integral Theory Conference*. Retrieved August (Vol. 1, p. 2010).
- O'Fallon, T. (2011). STAGES: Growing up is Waking up—Interpenetrating Quadrants, States and Structures. Available at www.pacificintegral.com.
- O'Fallon, T. (2012). Development and consciousness: Growing up is waking up. *Spanda Journal*, III(1), 97-103.
- O'Fallon, T. (2013, July). The senses: Demystifying awakening. Presented at the Integral Theory Conference.
- O'Fallon T., Ramirez, V., & Fitch, G. (2014). Collective intelligence as a causal ground. *Spanda Journal*, V(2), 91-95.
- Pals, J. L., & John, O. P. (1998). How are dimensions of adult personality related to ego development? An application of the typological approach. In P.M. Westenberg, A. Blasi, & L.D. Cohn (Eds.), *Personality development: Theoretical, empirical, and clinical investigations of Loevinger's conception of ego development* (pp. 113-131). Mahwah, NJ: Erlbaum.
- Pellegrino, J. W., & Hilton, M. L. (Eds.). (2012). Education for life and work: Developing transferable knowledge and skills in the 21st century. National Research Council Committee Report, National Academies Press.
- Pfaffenberger, A. H. (2005). Optimal adult development: An inquiry into the dynamics of growth. *Journal of Humanistic Psychology*, 45(3), 279-301.
- Pfaffenberger, A. H., Marko, P. W., & Combs, A. (Eds.). (2011). *The postconventional personality: Assessing, researching, and theorizing higher development*. SUNY Press.
- Ramirez, V., Fitch, G., & O'Fallon T. (2013). Causal leadership: A natural emergence from later stages of awareness. Paper presented at the Integral Theory Conference, San Francisco, CA.
- Redmore, C., & Waldman, K. (1975). Reliability of a sentence completion measure of ego development. *Journal of personality assessment*, 39(3), 236-243.
- Rifkin, J. (2009). *The empathic civilization: The race to global consciousness in a world in crisis*. Penguin.
- Rooke, D., & Torbert, W. R. (1998). Organizational transformation as a function of CEO's developmental stage. *Organization Development Journal*, 16(1), 11.
- Rooke, D., & Torbert, W. R. (2005). Seven transformations of leadership. *Harvard Business Review*, 83(4), 66-76.
- Rosenberg, S. W. (2004). Reconstructing the concept of democratic deliberation. Center for the Study of Democracy.
- Rosenberg, S. W. (2007). Rethinking democratic deliberation: The limits and potential of citizen participation. *Polity*, 39(3), 335-360.
- Scardamalia, M., Bransford, J., Kozma, R., & Quellmalz, E. (2012). New assessments and environments for knowledge building. Assessment and Learning of 21st Century Skills Bereiter, C., & Scardamalia, M. (2012). What Will It Mean To Be An Educated Person in Mid-21st Century?
- Sompolinsky, H. (1987). "The theory of neural networks: The Hebb rule and beyond." In *Heidelberg colloquium on glassy dynamics*, pp. 485-527. Springer Berlin/Heidelberg, 1987.
- Stein, Z. (2008). Myth busting and metric making: Refashioning the discourse about development. Excursus for Integral Leadership Review. *Integral Leadership Review*, 8(5).
- Stein, Z., & Heikkinen, K. (2009). Models, metrics, and measurement in developmental psychology. *Integral Review*, 5(1), 4-24.
- Thompson, W. I. (1998). *Coming into being: Artifacts and texts in the evolution of consciousness*. Palgrave Macmillan.
- Thompson, W. I. (2007). Natural drift and the evolution of culture. *Journal of Consciousness Studies*, 14(11), 96-116.
- Thoughts on Loevinger's theory in relation to subject-object psychology. In P. M. Westenberg, A. Blasi, & L. D. Lawrence (Eds.), *Personality development: Theoretical, empirical, and clinical investigations of Loevinger's conception of ego development* (pp. 39-58). Mahwah, NJ: Lawrence Erlbaum.
- Torbert, W. R. (2014). Brief comparison of five developmental measures: The GLP, the MAP, the LDP, the SOI, and the WUSCT primarily in terms of pragmatic and transformational validity and efficacy. Available from Action Inquiry Associates, www.williamrtorbert.com.
- Torbert, W. R., & Livne-Tarandach, R. (2009). Reliability and validity tests of the Harthill Leadership Development Profile in the context of Developmental Action Inquiry theory, practice and method. *Integral Review*, 5(2), 133-151.
- Vincent, N. C. (2014). Evolving consciousness in leaders: promoting late-stage conventional and post-conventional development (Doctoral dissertation)
- Vincent, N., Ward, L., & Denson, L. (2013). Personality preferences and their relationship to ego development in Australian leadership program participants. *Journal of Adult Development*, 20(4), 197-211.
- Vincent, N., Ward, L., & Denson, L. (2015). Promoting post-conventional consciousness in leaders: Australian community leadership programs. *The Leadership Quarterly*, 26(2), 238-253.
- Vincent, N., Ward, L., & Denson, L. (2015). Promoting post-conventional consciousness in leaders: Australian community leadership programs. *The Leadership Quarterly*, 26(2), 238-253.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, (Eds.). Cambridge, MA: Harvard University Press.

- Walsh, R. (2015). Wise Ways of Seeing: Wisdom and Perspectives. *Integral Review: A Transdisciplinary & Transcultural Journal for New Thought, Research, & Praxis*, 11(2).
- Westenberg, M. P., Drewes, M. J., Goedhart, A. W., Siebelink, B. M., & Treffers, P. D. (2004a). A developmental analysis of self-reported fears in late childhood through mid-adolescence: social-evaluative fears on the rise? *Journal of Child Psychology and Psychiatry*, 45(3), 481-495.
- Westenberg, P. M., & Block, J. (1993). Ego development and individual differences in personality. *Journal of Personality and Social Psychology*, 65(4), 792.
- Westenberg, P. M., Hauser, S. T., & Cohn, L. D. (2004). Sentence completion measurement of psychosocial maturity. *Comprehensive handbook of psychological assessment*, 2, 595-616.
- Wigglesworth, C. (2006). Why spiritual intelligence is essential to mature leadership. *Integral Leadership Review*, 6(3), 1-17.
- Wigglesworth, C. (2012). *SQ21: The twenty-one skills of spiritual intelligence*. BookBaby.
- Wilber, K. (1995). Sex, Ecology, Spirituality: The Spirit of Evolution. Boston and London: Shambhala.
- Wilber, K. (2000A). *Integral Psychology*: Consciousness, spirit, psychology, therapy. Boston, MA: Shambhala Publications.
- Wilber, K. (2005). *A sociable god: Toward a new understanding of religion*. Shambhala Publications.
- Wilber, K. (2006). *Integral Spirituality*. Boston, MA: Shambhala Press.
- Wilber, K. (2007). *A brief history of everything*. Shambhala Publications.
- Wilber, K., Patten, T., Leonard, A., & Morelli, M. (2008). *Integral life practice: A 21st-century blueprint for physical health, emotional balance, mental clarity, and spiritual awakening*. Shambhala Publications.

Appendix 1 — Summary of Conclusions in the Article

Here is an overview of the issues and conclusions in this article, mostly by Section title:

Background Context and Preface. In this section I discuss the larger context and motivations for this paper, which is to provide theoretical and empirical validity background information for STAGES; and to do this in support of an R&D project creating a computer-based AI scoring system for the SCT.

An introduction to wisdom skills and adult development. In the Introduction we argue for the importance of supporting wisdom skills to address contemporary society's complex issues. We draw from the literature on wisdom skills to say that they include: self-awareness, seeing big pictures and contexts, perspective-taking and empathy, tolerance for uncertainty and paradox, and pragmatic judgment that grounds ideation in observable reality. We argue that holistic developmental theories of adult development, including Loevinger's and Kegan's theories, successfully, if roughly, integrate this set of skills within a single construct (ego development or meaning making maturity). Thus, to a first approximation, the associated assessments are powerful indicators of wisdom skill. These skills essentially deal with the relationships among three realms: self, others, and the world. Unlike most scholarship looking explicitly at wisdom, these theories describe the *developmental* growth of these skills through levels of increasingly adequate and mature capacities. In the Introduction we give an overview of the basic principles of developmental theories, for those new to that topic.

Defining and measuring ego development. Next we move from giving a general appreciation of wisdom skill to focusing in on Loevinger's concept of ego development and its measurement —exploring the definition and range of the construct from an academic, research, and validity perspective. We describe the Washington University Sentence Completion Test (WUSCT) comprised of 36 sentence stems, along with its scoring manual and methodology. This is a "projective" developmental personality test, and is thought to capture a more accurate representation of both the conscious and unconscious aspects of maturity (compared with, for example, fixed-choice self-rating instruments). We describe the WUSCT and the method of calculating the "total protocol score," touching on the nuances of using "cut-off" values. We begin to discuss the complex meaning of "ego development" as it relates to constructs such as complexity intelligence, social/emotional intelligence, and personality traits, and discuss Loevinger's success in establishing a single construct (ego development) that integrates capacities across these domains.

Finally we describe how Cook-Greuter's research increased the understanding of the later (post-autonomous) developmental levels and added to the explanatory structure to the overall arc of development. Cook-Greuter frames development in terms of perspective-taking growth, and points out the deconstructive or emptying elements of the later stages of meaning-making.

Internal and External validity of SCTs. What does the research say about the validity of WUSCT and ego development theory, in terms of internal and external validity? The literature drawing on Loevinger's model is extensive, including at least 400 studies and at least a half dozen meta-analysis and critical overviews. This body of work compellingly supports the validity and usefulness of the SCT. We summarize the substantial body of work supporting the exceptional psychometric properties of the WUSCT in terms of:

- Internal Validity:
 - inter-rater reliability,
 - internal consistency,
 - test-retest reliability.
- External Validity:
 - face validity,
 - construct validity (concurrent, convergent, discriminant, and incremental validity), and
 - predictive validity.

We describe each type of validity criteria along the way. We discuss the sensitivity of the WUSCT to "retesting," and "faking, guessing, and scaffolding"—concluding that it is generally robust, though we note specific vulnerabilities.

In the discussion of construct validity we review what research studies say about the relationship between ego development and intelligence (IQ), verbal fluency, socioeconomic status, education level, socioeconomic status, moral judgment, leadership behaviors, antisocial behavior, adaptation after personal crises, and personality factors such as openness to experience and conformity. We also note its use in different countries, cultures, and languages.

Internal and External validity of modifications to the WUSCT. In Section 5, to anticipate the evolution of the WUSCT to the modified SCTs developed by Cook-Greuter, Torbert, O'Fallon, and colleagues, we explore the literature about how various changes to the WUSCT impact its validity. Based on the research literature, we conclude:

- Re **length** variations: shorter versions of the WUSCT (or any SCT), including half-length (18 item) versions, maintain strong reliability, though of course the longer the test the stronger its reliability.
- Re **minor stem** variations: Various attempts to use alternative sentence stems have not impacted the general reliability of the SCT. As in any item-based test, new stems must be tested for clarity and fit, but the general approach of the projective SCT seems to maintain its validity over different choices of stems. It is believed that a robust test should include stems from a suitable (or holistic) variety of life contexts to cover the same territory as the WUSCT. Most modifications do not change more than 6 or 8 stems out of the 36.
- Re **major stem** variations. As more sentence stems are modified, the scope or focus of meaning making or wisdom maturity is expected to shift somewhat, so that significant variations on the WUSCT are expected to maintain *validity*, but to be increasingly less *comparable* with each other. This is particularly relevant to O'Fallon's experimentation with "specialty protocols" that include stems focusing on particular areas (leadership, relationship, money, etc.).
- Re **assessment method**. The classic ego development test assumes particular instructions, which, so far all have kept constant. Moving from paper to on-line formats does not seem to appreciably affect outcomes. As researchers, such as O'Fallon, move toward assessing other types of text (such as essay answers), gathered using non-standard methods, ego development theory can be applied, but more research will be needed to demonstrate validity. We also mention alternative ways to aggregate across the 36 sentences to produce a "total score."
- Re **construct indeterminacy**. We also make the point that each variation in measurement method shifts the meaning of the measured construct (e.g. ego development) slightly, and discuss why there is no definitive answer to the question of which measurement or model gives the "correct" meaning to a construct.

MAP/GLP/LDP: internal and external validity. In Sections 5 & 6 we move from discussing the validity (5. internal validity and 6. external validity) of the WUSCT to its successor assessments. Cook-Greuter and Torbert developed the MAP/GLP/LDP instruments based on the WUSCT (we save O'Fallon's STAGES model for a later section). They changed the length and some of the stems of the

standard WUSCT (adding new sections to the scoring manual). They added two additional developmental levels to the framework and its scoring manual. They also moved from using the SCT only for research purposes, as Loevinger did, to using it as a commercial service that includes coaching/debriefing advice. Cook-Greuter, Torbert, and colleagues have accumulated a wealth of practical anecdotal evidence of the usefulness and face validity of their version of the SCT. In addition a handful of scientific studies of the MAP/GLP instruments have successfully demonstrated the following types of validity (details given in the later section): inter-rater reliability, internal consistency, construct validity, and predictive validity.

Regarding external and predictive validity of the newer SCTs, as noted in Vincent (2015) "a growing body of studies is showing associations between increasing consciousness development and better leadership performance and organizational outcomes." In studies of limited size, both the SCT and Kegan's SOI have been used to show strong relationships between: development and traits associated with mature leadership; 360-degree leadership feedback instruments, and leadership success in complex organizational transitions.

The STAGES Model. STAGES proposes that there is an underlying repeating structure that explains the developmental levels found by Loevinger and others. She borrows from Wilber's AQAL theory giving four principle dimensions, including three polarities from AQAL and an additional three-step Tier dimension. Development progresses through the Concrete, Subtle, and MetAware tiers; and within each tier it moves through two person-perspectives (Individual and Collective), and within each PP it passes through two stages (Passive and Active). A further refinement of Early and Late within each stage can be made using the Exterior/Interior dimension. In all this defines 12 states of development: Impulsive, Egocentric, Rule oriented, Conformist, Expert, Achiever, Pluralist, Strategist, Construct Aware, Transpersonal, Universal, and Illumined. O'Fallon also uses the terms Receptive, Active, Reciprocal, and Interpenetrative to describe the sequence of 4 levels within each tier.

O'Fallon builds upon Wilber's "Wilber-Combs Matrix" model to suggest that certain states seem to be necessary but not sufficient for certain stages. We use the principle from cognitive science that "neurons that wire together fire together; and those that fire together wire together" to explain the state vs. stage relationship.

We describe some striking similarities between the STAGES model and the Fischer's Skill Theory, an advanced neo-Piagetian theory of hierarchical complexity in development.

Finally we summarize a STAGES validity study indicating strong psychometric properties of the measurement and strong replicability vs. Cook-Greuter's ego development measure (and thus Loevinger's as well).

Appendix 2 — Developmental Level Correspondences

Below are table showing correspondences between developmental models, gleaned from various sources in the literature. Definitions of levels of different models do not usually line up in any exact way, but share enough characteristics that they are comparable. (Items with a question mark are more open to interpretation or placed in questionable locations.)

Most scholars agree that there is substantial overlap—that the various theories talking more or less about the same thing. But there are differences as well, and these become important as one tries to compare one theory to another, or an assessments made within one model to assessments made within another. Though there are differences, the similarities and alignments are actually quite striking, given that many of them were developed independently, which supports the overall face validity of the developmental approach.

Table 3: Correspondences of Levels in Some Developmental Models

O'Fallon	Loevinger	Torbert	Cook-Greuter	Kegan
----------	-----------	---------	--------------	-------

Illumined (6.5)				
Universal (6)		Ironist	Unitive	
Transpersonal (5.5)				
Construct Aware (5)	Integrated e9	Alchemist/ Magician	Construct-aware / Ego-aware	
Strategist (4.5)	Autonomous e8 (i5)	Strategist	Autonomous	Self transforming/ Inter-individual (5th)
Individualist (4)	Individualistic e7 (i4/5)	Individualist	Individualist	
Achiever (3.5)	Conscientious e6 (i4)	Achiever	Conscientious	Self authoring/ Institutional (4th)
Expert (Specialist) (3)	Self-aware e5 (i3/4)	Expert/ Technician	Self-conscious	
Diplomat (conformist) (2.5)	Conformist e4 (i3)	Diplomat	Conformist	Socialized/ Interpersonal (3rd)
Rule oriented (2)	(Delta/3)			
Opportunist (Ego centric) (1.5)	Self-protective e3 (i 2/3, opportunist, delta,)	Opportunist	Self-defensive	Instrumental/ Imperial (2nd)
Impulsive (1)	Impulsive e2 (i2)	Impulsive	Impulsive	Impulsive (1st Order)
	Pre-social e1			

Table 3 cont.: Correspondences of Developmental Model Levels -- others

O'Fallon	Wilber	Beck (Wilber colors)	Joiner	Commons	Fischer
Illumined (6.5)					
Universal (6)	Meta-Mind?	(Violet?)		Cross-parad.?	
Transpersonal (5.5)	Global mind?	(Indigo?)			
Construct Aware (5)	Late vision logic	Turquoise	Synergist	13. Para-digmatic	
Strategist (4.5)	Middle vision logic	Yellow—Integrated Self (<i>Teal</i>)	Co-creator	12. Meta-systematic	(P1/Ab4) Single Principles/ Axioms
Individualist (4)	Early vision logic	Green—Sensitive Self	Catalyst	11. Systematic	(Ab3) Abstract Systems
Achiever (3.5)	Formal operational	Orange—Achiever Self	Achiever	10. Formal	(Ab2) Abstract Mappings
Expert (Specialist) (3)			Expert	9. Abstract	(Ab1/Rp4) Single Abstractions
Diplomat (Conformist) (2.5)	Concrete operational	Blue—Traditional/ Rule-Role Self (<i>Amber</i>)	Conformer	8. Concrete	(Rp3) Representational Systems
Rule oriented (2)				7. Primary	(Rp2) Representational Mappings
Opportunist (Ego centric) (1.5)	Pre-operational (concrete)	Red—Impulsive Self	Operator	6. Pre-operational	(Rp1/A4) Single Representations
Impulsive (1)	Pre-operational (symbolic)	Purple —Magic/ Animistic Self	Enthusiast	5. Sentential	(A3?) Action (sensorimotor) Systems
	Sensorimotor	Beige—Instinctive/ Survival Self	Explorer		

O'Fallon has speculated a fourth Tier in the STAGES model, "Unitive," continuing the repeating patterns seen in the three Tiers in Table 2: 7.0, 7.5, 8.0, 8.5. This fourth Tier is based on her reading of Auobino's description of these rare stages of development. She puts Wilber's "Supermind" at 8.0.

Appendix 3 — Descriptions of Developmental Levels

Below we show several variations on how others summarize adult developmental levels. It is not the purpose of this article to describe developmental levels in any detail, nor to compare and contrast descriptions of levels given in different theories. We offer descriptions below to give the reader unfamiliar with this field a general sense of what changes as individuals develop through these levels.

These focus mostly on earlier and mid levels, and see more details on characteristics of post-autonomous (or post-rational) developmental stages in the subsection "Cook-Greuter's study of Post-autonomous levels."

First are two short descriptions of the progression of levels, followed by some tabular descriptions.

(See also, "Table 2.3: Stages of Consciousness Development" from Vincent, 2014.)

1. Short Summaries of Developmental progressions

Vincent (2015, p. 238):

"As meaning-making expands, so does self- and interpersonal awareness, cognitive complexity, flexibility, reflective capacity, personal autonomy and responsibility, as well as tolerance for difference and ambiguity."

Pals & John (1998, p. 120) summarizes a number of studies of empirical relationships between ego development and personality trait constructs.

Some of the relationships were linear, increasing monotonically with increases in ego development. In particular, the higher individuals are in ego development, the more they are open to experience, appreciate alternative values and new ideas, and have aesthetic interests. They are more creative and psychologically minded and think in more complex ways. Finally they have more mature, mutually respectful relationships, adopt more androgynous gender role definitions, and they are more responsive to psychological conflict. Other relationships we're not linear but Peak at middle or moderate levels of ego development. Thus, individuals at middle levels are the most compliant, interpersonally pleasant, and authoritarian. They are more likely to tightly control their needs and feelings and closely follow gender norms. At the same time, they feel the most emotionally secure. Finally, some traits seem characteristic of only the small subset of adults at the very low end of ego development: they show signs of maladjustment, specially hostile impulsivity, fear, and anger.

Barker & Torbert (2011, p. 55) report on Nicolaides (2008) on ambiguity:

...unlike people at conventional action logics who tend to try to avoid ambiguity, all of her post conventional sample saw positive, creative potential in ambiguity. But within this broad similarity, she found for distinctive responses to ambiguity: the Individualist...endured it; the Strategist...tolerated it; the Alchemists...surrendered to it, and the Ironist...generated it.

2. Manners 2001, p. 544:

TABLE 1: Loevinger's Stages of Ego Development

Presocial and Symbiotic (E1)	Exclusive focus on gratification of immediate needs; strong attachment to
---------------------------------	---

	mother, and differentiating her from the rest of the environment, but not her/himself from mother; preverbal, hence inaccessible to assessment via the sentence completion method.
Impulsive (E2)	Demanding; impulsive; conceptually confused; concerned with bodily feelings, especially sexual and aggressive; no sense of psychological causation; dependent; good and bad seen in terms of how it affects the self; dichotomous good/bad, nice/mean.
Self-Protective (E3)	Wary; complaining; exploitative; hedonistic; preoccupied with staying out of trouble, not getting caught; learning about rules and self control; externalizing blame.
Conformist (E4)	Conventional; moralistic; sentimental; rule-bound; stereotyped; need for belonging; superficial niceness; behavior of self and others seen in terms of externals; feelings only understood at banal level; conceptually simple, "black and white" thinking.
Self-Aware (E5)	Increased, although still limited, self-awareness and appreciation of multiple possibilities in situations; self-critical; emerging rudimentary awareness of inner feelings of self and others; banal level reflections on life issues: God, death, relationships, health.
Conscientious (E6)	Self evaluated standards; reflective; responsible; empathic; long term goals and ideals; true conceptual complexity displayed and perceived; can see the broader perspective and can discern patterns; principled morality; rich and differentiated inner life; mutuality in relationships; self critical; values achievement.
Individualistic (E7)	Heightened sense of individuality; concern about emotional dependence; tolerant of self and others; incipient awareness of inner conflicts and personal paradoxes, without a sense of resolution or integration; values relationships over achievement; vivid and unique way of expressing self.
Autonomous (E8)	Capacity to face and cope with inner conflicts; high tolerance for ambiguity and can see conflict as an expression of the multifaceted nature of people and life in general; respectful of the autonomy of the self and others; relationships seen as interdependent rather than dependent/ independent; concerned with self-actualization; recognizes the systemic nature of relationships; cherishes individuality and uniqueness; vivid expression of feelings.
Integrated (E9)	Wise; broadly empathic; full sense of identity; able to reconcile inner conflicts, and integrate paradoxes. Similar to Maslow's description of the "self-actualized" person, who is growth motivated, seeking to actualize potential capacities, to understand her/his intrinsic nature, and to achieve integration and synergy within the self (Maslow, 1962).

3. Loevinger's Ego Development Levels

From *Ego Development: Conceptions and Theories*, (p. 24-25) by J. Loevinger and A. Blasi, 1976, San Francisco: Jossey-Bass. (Also found in Vincent, 2014; and see also Loevinger 1998, p 9. "Some Characteristic of Level of Ego Development.)

STAGE	IMPULSE CONTROL, CHARACTER DEVELOPMENT	INTERPERSONAL STYLE	CONSCIOUS PREOCCUPATION	COGNITIVE COMPLEXITY
Impulsive	Impulsive, fear of retaliation	Receiving, dependent, exploitative	Bodily feelings, especially sexual	Stereotyping, conceptual confusion
Self- Protective	Fear of being caught, externalizing blame, opportunistic	Wary, manipulative, exploitative	Self-protection, trouble, wishes, things, advantage, control	
Conformist	Conforming to external rules, shame, guilt for breaking rules	Belonging, superficial niceness	Appearance, social acceptability, banal feelings, behavior	Conceptual simplicity, stereotypes, clichés
Self-Aware	Differentiation of norms, goals	Aware of self in relation to group, helping	Adjustment problems, reasons, opportunities (vague)	Multiplicity
Conscientious	Self-evaluated standards, self-criticism, guilt for consequences, long-term goals and ideals	Inventive, responsible, mutual, concern for communication	Differentiated feelings, motives for behavior, self-respect, achievements, traits, expression	Conceptual complexity, ideas of patterning
Individualistic	<i>Add:</i> Respect for individuality	<i>Add:</i> Dependence as an emotional problem	<i>Add:</i> Development, social problems, differentiation of inner life from outer	<i>Add:</i> Distinction of process and outcome
Autonomous	<i>Add:</i> Coping with conflicting inner needs, toleration	<i>Add:</i> Respect for autonomy, interdependence	Vividly conveyed feelings, integration of psychological causation of behavior, role conception, self-fulfillment, self in social context	Increased conceptual complexity, complex patterns, toleration of ambiguity, broad scope, objectivity
Integrated	<i>Add:</i> Reconciling inner conflicts, renunciation of unattainable	<i>Add:</i> Cherishing of individuality	<i>Add:</i> Identity	

NOTE: "Add" means in addition to the description applying to the previous level.

4. Cook-Greuter (2004, p. 7)—Some examples of how different action logics matter

First, let's look at how someone's understanding and response to the **concept of "feedback"** changes with increasing development.

Magician	View feedback (loops) as a natural part of living systems; essential for learning and change; and take it with a grain of salt.
Strategist	Invite feedback for self-actualization; conflict seen as an inevitable aspect of viable and multiple relationships
Individualist	Welcome feedback as necessary for self-knowledge and to uncover hidden aspects of their own behavior
Achiever	Accept feedback especially if it helps them to achieve their goals and to improve

Expert	Take it personally, defend own position; dismiss feedback from those who are not seen as experts in the same field (general manager)
Diplomat	Receive feedback as disapproval, or as a reminder of norms
Opportunist	React to feedback as an attack or threat.

Let's now look at what **methods of influence** people at different stages might use.

Magician	Reframe, turn inside-out, upside-down; clowning; holding up mirror to society; often behind the scenes.
Strategist	Lead in reframing, reinterpreting situation so that decisions support overall principle, strategy, integrity and foresight
Individualist	Adapt (ignore) rules where needed; or invent new ones; discuss issues and air differences
Achiever	Provide logical argument, data, experience; make task/goal-oriented contractual agreements
Expert	Give personal attention to detail and seek perfection; argue own position and dismiss others' concerns.
Diplomat	Enforce existing social norms; encourage, cajole; require conformity to protocol to get others to follow.
Opportunist	Take matters into own hands, coerce, win fight

Cook-Greuter, 2004, p. 279—Ego Development Stages
[Duplicated from Table 1 in the article]

Stage/Action Logic	Main focus	% adult pop. N=4510
Alchemist and above Deep processes and intersystemic evolution rules principles	Interplay of awareness, thought, action, and effects; transforming self and others	2.0
Strategist Most valuable principles rule relativism	Linking theory and principles with practice, dynamic systems interactions	4.9
Individualist Relativism rules single system logic	Self in relationship to system; interaction with system	11.3
Achiever System effectiveness rules craft logic	Delivery of results, effectiveness, goals, success within system	29.7
Expert Craft logic rules norms	Expertise, procedure and efficiency	36.5
Diplomat Norms rule needs	Socially expected behavior, approval	11.3
Opportunist and below Needs rule impulses	Own immediate needs, opportunities, self-protection	4.3

5. Table 1 from McCauley et al. 2006: Comparison of three constructive-developmental frameworks

Table 1
Comparison of three constructive-developmental frameworks

Framework	Dependent		Independent		Inter-independent	
1) <i>Kegan's Orders of Consciousness</i>	<i>Interpersonal/Traditional</i>		<i>Institutional/Modern</i>		<i>Interindividual/Post-modern</i>	
What is object?	Enduring needs and dispositions		Interpersonal relationships		The autonomous self	
What is subject?	Interpersonal relationships		The autonomous self		The transforming self	
2) <i>Torbert's Stages</i>	<i>Diplomat</i>	<i>Expert</i>	<i>Achiever</i>	<i>Individualist</i>	<i>Strategist</i>	<i>Alchemist</i>
Action logic	Norms rule needs	Craft logic rules norms	System effectiveness rules craft logic	Relativism rules single system	Most valuable principles rule relativism	Deep processes and intersystemic evolution rule principles
Main focus	Socially expected behavior, approval	Expertise, procedure, and efficiency	Delivery of results, effectiveness, success within system	Self in relationship to system; interaction with system	Linking theory and principles with practice, dynamic systems interactions	Interplay of awareness, thought, action, and effects; transforming self and others
3) <i>Kohlberg's Stages of Cognitive Moral Development</i>	<i>Mutual expectations</i>	<i>Law and order</i>	<i>Social contracts</i>	<i>Universal principles</i>		
Motives for moral action	Anticipation of disapproval from others	Anticipation of failure of duty	Concern to maintain self-respect and respect of the community	Concern about self-condemnation for violating one's own principles		
Definition of "right"	Being concerned about other people and their feelings; Being motivated to follow rules and expectations	Upholding social order and maintaining the welfare of the society or group	Upholding the basic rights, values, and legal contracts of society	Guidance by universal ethical principles that all humanity should follow		

Adapted from Cook-Greuter (2004), Kegan (1994), and Snell (1996).

6. Description of the STAGES levels and person-perspectives (PP), adapted from Fitch, 2016.

See Figure 2 and Table 2 in the STAGES Section above for illustrations.

1st PP: In the first person perspective, we are in the concrete 'I' stages. (In the STAGES model, these are called 1.0 Impulsive and 1.5 Opportunist). It is 'all about me' and there is no understanding yet of a 'We.' One can see others but does not have a truly unique identity separate from others, nor does one see others as unique in their own right. The focus is on one's concrete needs and wants.

2nd PP: The second person perspective foregrounds the concrete 'We' (2.0 Rule-oriented and 2.5 Conformist). In this perspective, I see that others see me and that, in order to satisfy my needs, I must work with others and make and follow rules together. In these 'We' stages, the 'I' is present and understood, but backgrounded, or deprioritized, in favor of relationships and groups.

3rd: The next perspective gives rise to subtle 'I' stages, where we realize we have a subtle self – the thoughts, emotions, and independent mind of rational consciousness (3.0 Expert and 3.5 Achiever in the model). Again this is an I-oriented space, while the we is present and backgrounded. The 'We' that is present, however, is the concrete we, groups and their norms and rituals, since no new subtle 'We' has yet been discovered.

4th PP: This pattern continues with the fourth person perspective, where the subtle 'We' is foregrounded. The subtle We consists of the perception of one being situated in and arising out of a plurality of contexts (4.0 Pluralist and 4.5 Strategist stages). The 'We' isn't then a specific group, but it is a space, and that space is complex. It consists not only of outer manifestations, such as the room, the systems in which the context is embedded, the cultural context and form, but also inner manifestations, such as the attitudes, beliefs, assumptions, states of awareness, and ontological dispositions we bring to the moment. Here, people begin to understand that one can have, if the context is right, an experience of the kind of deep connection to another, once thought reserved for a soul mate. It is at this level that the notions of we-space, collective intelligence, and collective evolution begin to arise, although we can say that collective intelligence takes form at concrete, subtle and Metaware levels.

5th PP: At the fifth person perspective, individuals awaken to their ever-present awareness as the ground of their own being. They begin to identify with this being as a new self, both empty and full, transcendent and immanent. In these stages (5.0 Construct Aware and 5.5 Transpersonal), the 'I' is foregrounded but the subtle 'We' remains as a context for this I. The 'I' is however not what we conventionally think of as 'I' – our concrete bodily self or our subtle thinking or narrative self, but rather our causal self, the limitless open horizon of awareness that we paradoxically seem to share with everyone and everything. Knowledgeable that deep subtle we-spaces are possible, individuals often desire to experience them in this causal experience, for example, by bringing the practice of witness consciousness into their collective experience, and to provide a context in which their causal selves can express themselves and be recognized.

6th PP: At the sixth person perspective, this new 'I' is again backgrounded as it lets go into a much larger, causal 'We.' In the sixth level stages (6.0 Kosmic and 6.5 Illumined), the 'We' is all of concrete, subtle, and causal manifestation itself, the Kosmos, the utterly full and empty existence, eternal and beyond time, infinite and beyond space. Here one experiences themselves as this whole, with their apparent (even causal) 'I' birthed by and birthing the whole. There is a keen interest at the sixth level in living as this larger collective, which has its own sense of 'We', and in allowing the intelligence of the whole, and that which births the whole, to express one's existence. In this sense, this very much reflects the interest in collective intelligence at the fourth stage, but is no longer grounded in any specific group or context but in reality itself.

[STAGES uses the Passive/Active parameter (or, equivalently, the Receptive/Active/Reciprocal /Interpenetrative terminology) to further differentiate the 6 Person Perspectives into all 12 level (with X.0 and X.5), which have rich descriptions that correspond to the Loevinger/Cook-Greuter/Torbert levels, and also roughly to Kegan's levels.]

Page of

[1] <http://integralleadershipreview.com/table-of-contents/?slug=august-november-2017>

[2] The original purpose was to provide background information for a related project—the author is developing an automated assessment tool for scoring the types of sentence completion tests (SCT) described in the article (information at OpenWaySolutions.com).

[3] See OpenWaySolutions.com

[4] With large sample sizes it is also possible to use much shorter versions of the 36-item sentence completion test, helping participant recruitment by lowering the effort required to participate. We are also researching the analysis of arbitrary text (beyond sentence completions) for developmental level, which further reduced the barriers to gathering data for developmental research and analysis.

[5] Attaining more advanced levels of development is not an end or goal in itself, and can actually be maladaptive. Supporting development is important when the complexity demands of one's situation (work, relationships, etc.) outstrip one's capacities.

[6] It complicates matters that what are considered characteristics of wisdom may vary according to developmental level, but here we make assumptions that the reader has "post-conventional" meaning-making.

[7] Our list of skills also has significant overlap with Wigglesworth's description of "spiritual intelligence," which she breaks down into 21 skills distributed over Wilber's four quadrants, I/we/it/its. In fact Wigglesworth's measurement of spiritual intelligence shows a high correlation with ego development scores (Wigglesworth 2006, 2012).

[8] E.g. Cook-Greuter (2013, p. 2): "Development theory describes the unfolding of human potential towards deeper understanding, *wisdom* and effectiveness in the world" (my emphasis). Each term has its limitations and awkwardness—"wisdom skill" may seem as pretentious; "ego development" as technical jargon; and "meaning making" as too vague. The term "perspective taking" ability is sometimes used, but has an overly specific and constrained meaning in the research literature.

[9] Progressive theories of adult learning and development, including the "transformational learning" theory of Meizrow (1991, 2000) and "self directed learning" the theory of Knowles (1970) are only weakly related to the "construct-developmental" and theories we cover. These theories frame adult learning in terms of encountering disorienting dilemmas, self-reflection on assumptions, generative dialogue with others, creative inquiry into alternatives, iteratively putting alternatives to the test, and even examining ones' deeper meaning making foundations (motives, values, beliefs, behaviors, and identities). For more on this see Murray, 2009 "What is the Integral in Integral Education? From progressive pedagogy to integral pedagogy."

[10] There is some disagreement or uncertainty as to whether development is more continuous or occurs in growth spurts—we discuss this more later. At the minimum, slicing a developmental trajectory into discrete levels provides a simplification that helps us understand and dialogue about the phenomena.

[11] There is some controversy in the literature around whether development happens continuously or as discrete levels in spurts. Research studies have shown both to occur, and it seems as though for more specific and independent skills levels do appear, while for complex skills, while each of the invisible underlying parts may be growing through its own sequential timing in spurts, the complex combined behavior on the surface appears more continuous (and see Dawson-Tunik et al., 2005). The consensus seems to be that the spurts, or "gappiness" that give learning a discrete level-to-level structure, or an "overlapping waves" pattern, rather than a continuous linear change pattern, are related to the "punctuated equilibrium" that systems science has identified in all domains of phenomena, from the structure of matter and galaxies to, the progression of species evolution on the earth, to conceptual learning (Gould & Eldredge, 1977).

[12] Westberger, 2004, uses the terms "characteristic vs. maximum" performance to describe the same phenomena.

[13] Cook-Greuter (2013, p. 3): "Derailment in development, pockets of lack of integration, trauma and psychopathology are seen at all levels. Thus later stages are not more adjusted or 'happier.'"

[14] As Blumentritt (2011, p. 164) puts it, "[ego] development is not (and has not been) related to happiness, subjective well-being, or psychological adjustment." An increase in the complexity of meaning-making can decrease or increase one's suffering, depending on many factors, including the fit of one's perspective with one's environment.

[15] Because sentence completion is free-response and interviews probe for deeper understanding, it is thought that the interview method usually yields a slightly higher level. (Kegan et al., 1998).

[16] Likewise Fischer & Pruyne (2002) illustrate how the construct and theory of "reflective thinking" has significant overlap with several others, including argumentative thinking, mindfulness, constructed knowing, mental systematicity, willingness to suspend judgment, and tolerance for cognitive dissonance.

[17] Manners & Durkin (2001, p. 550) show evidence that ego development is distinct from purely cognitive skill, and that "a highly significant relation was found when cognitive level was assessed in terms of its application to the *socioemotional domain*" (emphasis mine).

[18] And Miniard, (2009, p. 562) says that "The findings indicate that ego development may be regarded as a complex, but unitary, construct, with the ego developing in a hierarchical, invariant, and sequential manner."

[19] However, as we work on group-level assessment, as in the StageLens auto-scoring project, the emphasis is moving away from assigning a center of gravity to any group or subgroup, and rather to compare whole distributions.

[20] Based on the idea that, as in medical diagnosis, the rarer a symptom, the greater its predictive power.

[21] Though Loevinger defends the cutoff method, its limitations have been noted. Holt (1980, p. 918), comments on the reliability of using cutoff methods: "The ogive rules are ingenious procrustean devices to make gradual transitions look like phase changes, translating continuity into discontinuous types. Naturally, therefore, at the borderlines, scoring gets much more unreliable.... Furthermore, since these rules put special emphasis on the few most extreme scores, TPRs become even less reliable the farther we get from the center of the distribution of the population."

[22] Note that the research results given are meant to argue in general for the validity of the method, or illustrate areas of concern, and that results of different studies reported here can not be directly compared (without further explanation) because they may occur in different contexts (e.g. use in adult vs. teen populations; on-line administration vs. paper administration of the test; moderately trained vs. highly trained scorers, etc.)

[23] Internal validity is a prerequisite for, or contributes to, external validity, since a flawed measurement is less likely to accurately measure real phenomena. Validity can apply to theories, experiments, and measurements. We are concerned with measurement here, though all three are interlinked. The definitions of validity terminology vary by field and sometimes within field. The definitions we are using fit well for assessing a psychometric measurement by itself, as opposed to an experimental study per se.

[24] For example, using a person's height to measure how hard they are to carry may show good internal validity, i.e. the measurement is reliable and repeatable, but its external validity is not particularly good--better to use weight, even if all you have is an old rusty scale that is a bit unreliable (internal validity), you are still more likely to measure what you are wanting to measure (external validity).

[25] Correlations are generally thought to be inferior metrics for the ordinal TPR value.

[26] Even "self-trained" raters, who studied Loevinger's instructions but had no supervision, reached adequate IRR above 80% (vs. highly trained scorers: correlations of .76-.85 and agreements from 77% to 82%; Loevinger et al., 1970, p. 42-43).

[27] Holt (1980, p. 913; using test with fewer items) reports per-item agreements averaging about 79% (ranging from 66%-91%) and correlations of about .80.

[28] A less frequently reported statistic is the correlation between each stem and the whole profile (TPR), which Torbert & Livne-Tarandach (2009) report as ranging from .47 to .63 over the stems, with an average of .55.

[29] For any test we expect that a person's general stage of mind when they take the assessment (whether they are tired, emotionally upset,

distracted, etc.) will affect outcomes somewhat, leading to one source of unpredictable variation. However, for most contemporary (internet-delivered) applications of the SCT this effect is minimized because participants are able to take the test at a time that works for them, and there is (presumably) no pressure to excel.

[30] Loevinger also says (IBID p. 289) "Redmore and Waldman (1975) report several small studies of short-term retest effects...There appears to be a tendency for scores to decline on retest after a week or two, averaging about half a stage. This conclusion may apply particularly to scores above the Conformist Stage, since the elaborations and qualifications that establish a response as being at high levels are what the bored subject is apt to omit on retest." She reports on a study of short-term retesting where 60% scored the same, and of the 40% who changed all were within one level, evenly distributed between increasing and decreasing."

[31] For example King & Kitchener (2004) report on the assessment of reflective judgment using both a dilemma-based instrument (the RJI, a "production" task) and a fixed-choice instrument (the RCI, a "recognition" task). Correlations between the two were rather low (in the low 40's percent), with the recognition task showing higher scores than the production task.

[32] Westenberg (2004b, p. 603) claims that "Several studies suggest that the split-half reliability of the WUSCT...is about .80, and, if disattenuated for the greater unreliability of the two test halves, the correlation between the two halves approached unity."

[33] Browning (1987) used a short, 12-item, form in her study, and found that the short form has interrater agreement and internal consistency (alpha) "comparable to those reported by Loevinger" (p. 114).

[34] Holt also reports that the Cronbach's alpha for the 12-item test was .77, vs. Loevinger's finding for .91 for all 36 items, which is exactly what psychometric theory would predict for a randomly selected subset of 12 items from 36 (p. 915).

[35] Cohn and Westenberg (2004) "conducted a meta-analysis of 52 correlations between ego level scores and intelligence [and found] an average correlation between ego level and intelligence ranged from .20 to .34." and an overall weighted Pearson correlation of 0.30.

[36] Stein (2008, p. 15): "Developmental metrics are simply attempts to improve upon the ways we have always already made development judgments of each other and ourselves." Manners & Durkin, 2001, p. 543): "evaluating the theoretical and logical coherence of ego development is inseparable from evaluating the construct validity of the WUSCT."

[37] King (2011, p. 165) notes "Ego development is a difficult concept to grasp. Scholars have difficulty defining it. Once defined, it remains only slightly less inscrutable. Manuscript reviewers have a hard time 'getting it.' Surely people cannot consciously strive toward a developmental end that is so elusive and potentially baffling." Cohn & Westenberg (2004, p. 483) says that "for Loevinger the 'ego' is an abstraction, not an extant structure; thus, she describes the ego informally, referring to it as 'a frame of reference' or 'lens' through which individuals perceive their world (ego *development* thus represents a change in one's frame of reference)."

[38] This model has been applied by others extensively, to the analysis of dozens of domains (see Esbjörn-Hargens, 2009), and to interdisciplinary projects to connect domains normally separate within academic silos. One of the principles of Integral Theory is that all objects contain or enact all of the quadrants (zones) holistically, so the model can be used as a counter to reductionist methods of analysis.

[39] Note that STAGES is more of an application of AQAL elements than a direct extension of AQAL theory. AQAL categorizes all reality in one map, and developmental lines happen within each quadrant (and radiate out into all quadrants simultaneously). STAGES flips this ordering, and puts the map (repeated three times) *within* a developmental sequence (i.e. the tiers). This is discussed at more length in the Response article in this issue, where we note that Wilber's definition of "address = altitude + perspective," is turned around in STAGES to be "perspective = address + altitude"—where the terms take on a somewhat different meaning in the modified formula. STAGES is not an alternative to strict confirmation of AQAL because it is not an ontological theory about reality as a whole, but is specifically about human meaning making (epistemological or psychological reality). It supports AQAL in showing that the dimensions of reality that Wilber proposes as the most fundamental are the same dimensions shown empirically to fundamentally explain development in STAGES.

[40] In this sense a first-person/inside or third-person/outside perspective can be taken on any object; and this not to be confused here with the person-perspectives of the development levels, which go from 1st, 2nd, 3rd, 4th, 5th... Person Perspectives. The terminology is a bit confusing here, but basically, 3rd Person perspective is the level at which a person can begin to take a *formal* third person perspective that characterizes the modern or scientific objective viewpoint.

[41] We can note one point that may be confusing for readers familiar with AQAL who are learning about STAGES. In AQAL the Interior/Exterior dimension is one of the two primary dimensions (along with Individual/Collective). Inside/Outside (Passive/Active in STAGES) is introduced later through the 8-zones model. In STAGES, Interior/Exterior is the least important, and is often skipped over. Because of this, any diagram that tries to illustrate the STAGES progression layered on top of the typical AQAL zones diagram looks a bit tangled.

[42] In so doing one gains a deeper (though always limited) awareness of the contours of one's own unconscious. Just as a Subtle understanding of others includes a perception that underneath our concrete differences all people share a basic humanity; a Metaware understanding includes a perception of what all life, or even all matter or all reality, have in common at fundamental level one might simply call "being" or "awareness." In inhabiting the Metaware tier, whether routinely or momentarily, one experiences a heightened sense of connectedness, unity, wholeness, uniqueness, fullness, and emptiness. Note that this does not guarantee sustained happiness (or even morality) as one is also more open to the depth of suffering within the self, others, humanity, and the global ecosystem. Also, unresolved shadow material can within highly developed individuals be particularly pernicious.

[43] We are ignoring here that all individual things are, if one looks more closely, also collection of constituent things—as explained in the theory of holons (Wilber 200; Koestler, 1967). Individual holons are tightly coordinated compositions of highly differentiated parts acting as one object (the animal composed of many types of cells), while collective holons are more loosely connected collections of similar objects (the flock of birds) (and see Wilber's 10 Tenets of holons, Wilber, 2000). Borrowing from Mervis & Rosch's (1981) theory of concept development, we can say

that the first things we learn about are "natural types" at the most salient physical scale. We are aware of trees before we are aware of forest or leaves (and later, cells); and we are aware of people before we are aware of families, nationalities, or ears, noses, pupils, and (later) bladders and neurons. But as mentioned, to understand a collective is not only to simply see an object, but to be aware of the relationships among the objects such that one perceives an emergent whole.

[44] One also begins to understand the self as a collection of drives, beliefs, habits, and sub-personalities, which allows greater insight into the phenomena of projection and introjection.

[45] Note that, for example at 3.0, where one is in a receptive phase with respect to subtle individual objects, one continues in an active (including "interpenetrative") relationship to concrete objects.

[46] The precise understanding of each of these concepts for the purpose of scoring is non-trivial, and when there is ambiguity other factors are brought into scoring decisions, as explained in the STAGES scoring manual.

[47] Cook-Greuter's method of scoring the two latest stages that she added to the Loevinger model went one step in the direction of more generality in having a single set of exemplars and principles that held for all 36 stems—vs the scoring instructions for all earlier levels which have separate exemplars for each of the 36 stems.

[48] That is, rather than first observing that a set of protocols within one level seem to fall within *two* sequentially related categories, and then trying to articulate the definition of those two categories, as in Loevinger's method; the STAGES model itself defines additional subcategories, and the definitions of those sub-categories were used to subdivide the one prior category into two (for the three categories that were subdivided).

[49] Cook-Greuter's does propose some overarching structural themes. In *Nine Levels of Increasing Embrace in Ego Development* (2013). On page 15 she proposes, similar to the Spiral Dynamics model (Beck & Cowan, 1996), that the levels alternate between focusing on integration vs. differentiation. On page 16 she illustrates an "arc of increasing integration and differentiation" that peaks at between achiever and individualist levels.

[50] O'Fallon has created "specialty protocols" including a subset of stems targeted toward specific domains (so far these all have only 6 of the 36 stems changed). Several of these have been validated in the sense that enough data has been gathered to show that they have psychometric properties comparable to the original SCT (primarily the Cronbach's Alpha measure of internal consistency). Validated specialty protocol topics include: business, love, and education. In-development specialty protocols include coaching, parenting, and faith-based spirituality.

[51] Citations suggesting that Wilber may agree with O'Fallon's interpretation are given under "Issue #3" in the Response article in the s IRL issue.

[52] Neural wiring is determined by many things, not only learning and repetition; for example, innate genetically driven directives. And what actually fires in a given moment depends on the context of current and past stimulus, as well as on wiring.

[53] Repeated firing builds wiring (both excitatory and inhibitory connections). The biology is not completely understood yet, but repeated co-occurrence of stimuli or behavior can create new associations between previously unconnected nodes (not just strengthen weak connections).

[54] Though Loevinger, researching ego-development, and Kegan, researching "constructive developmental theory," consider Jean Piaget's work to be foundational for them, "Neo-Piagetian" work stems more directly from Piaget's theory and methodology.

[55] Appendix 1 shows correspondences between STAGES and Skill Theory levels.

[56] Those working with Common's and Fischer's theories make strong use of Rasch modeling, in which "Estimates of the position of persons, items, and response-scale steps are generated to optimize the fit of the data to the model" [not the other way around as in Loevinger's work] (Al-Owidha et al., 2009).

[57] There is emerging evidence (e.g. Aiden Thornton's as-yet-unpublished dissertation research comparing MAP and Lectica assessments) that ego development measurement (meaning-making maturity) does not correlate as strongly with Hierarchical Complexity (HC) assessments (such as reflective reasoning) as the integral community tends to assume. This would indicate that these two measurement traditions are looking at different phenomena. Some factors that may contribute to this difference include: (1) the SCT is a projective instrument that, arguably, can provide evidence about *non-conscious* social/emotional/ego thought processes, while the HC assessments tend to focus more on specifically cognitive skills, including one's *analytical understanding* of social/emotional/ego concepts; and (2) HC assessments use a strict model of increasing complexity through levels where each level builds upon prior levels, while ego development assessments include changes that may be in subtractive, ablative, or healing directions (and see Murray, 2009b).

[58] The method used was to cross-compare all combinations of four raters for each of the two scoring methods—a level of rigor that is rare, and perhaps unique, among the Loevinger-tradition research.

[59] The weighted Cohen's Kappa statistic for ordinal values was used to measure both replicability with CG/L and IRR for STAGES itself. It is the most often recommended measurement for inter-rating and replicability studies, and is a more stringent statistic than correlation or accuracy statistics, in part because it takes into account the possibility of random matches. Landis & Koch (1977) describes one of the most common systems for interpreting kappa magnitudes, describing the range 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1 as "almost perfect" agreement. Another often cited interpretation is that of Fleiss (1981), which describes values over .75 as "excellent."

[60] In the research summary given in O'Fallon (2013) the phrase "near perfect agreement" was used in alignment with convention. This precipitated words of critique from Cook-Greuter, which are understandable for those not familiar with the kappa statistic. Henceforward O'Fallon restrained from using those terms in favor of "very strong" or "excellent" for values higher than the cutoff for "substantial."