

8/31 – A Response to Critiques of the STAGES Developmental Model

2017-08-31 14:08:20 Terri O'Fallon, Tom Murray, Geoff Fitch, Kim Barta, and John Kesler

Terri O'Fallon, Tom Murray, Geoff Fitch, Kim Barta, and John Kesler

Executive Summary

Our goal in this article is to respond to the Critique's call for rigor by providing detailed explanations, and also, to provide further answers to questions and misconceptions that may exist about the STAGES model—therefore the article is rather long. For those who want a quick summary of how we respond to the Critique by Cook-Greuter, Wilber, and Sharma, we offer the following overview:

- **Preface and History.**

We appreciate the authors of the Critique for their remarkable contributions to the field and to us as individuals, and for giving us this opportunity to respond. We then describe historical events leading up to the current situation.

- **Acknowledgement of Issue #10.**

We first acknowledge the one concern we think was completely valid, at least in the past. Indeed, O'Fallon was making verbal and written statements about the statistical validity of her model without having a research report available that would allow colleagues to scrutinize those claims. Once her empirical study was complete and the statisticians began reporting positive results, O'Fallon and colleagues began to publicly mention the positive results. However, the publication of these results has taken much longer than expected, for reasons we explain (a peer review draft of the study is now near completion). Importantly, upon hearing Cook-Greuter's concerns about one year ago, O'Fallon and colleagues agreed with the concern and *did* remove mentions of the scientific validity of STAGES from their web site and other material. Thus, this concern is somewhat outdated.

- **Issue #1: STAGES “asserts specific descriptors based on a meta-model.”**

We clarify that STAGES is not a “theory of everything” as is AQAL, it is a model about the development of human meaning-making and perspective-taking that is based on ontological dimensions borrowed from AQAL. We describe similarities and difference between the dimensions of AQAL and STAGES.

- **Issue #2: STAGES “uses problematic terms to represent concepts.”**

The tiers in the STAGES model were originally called Gross, Subtle, and Causal, but, over one year ago, after hearing Wilber's concerns about the use of these terms, especially causal, O'Fallon changed the terms to Concrete, Subtle, and MetAware. Thus, this concern is somewhat outdated.

- **Issue #3: STAGES “conflates stages and states.”**

O'Fallon agrees with the basic tenets of the Wilber-Combs Matrix, but also proposes that repeated experience of certain states may be necessary (but not sufficient) for attaining certain stages. We cite instances where Wilber seems to agree with this. In addition, we propose an expanded description of what states and stages are that is grounded more in cognitive science. I.E. that states are about how neurons or nodes *fire*, and stages are about how neurons or nodes are *wired*. Using this definition, the relationship between states and stages can be taken from standard cognitive and brain science: nodes that fire together tend to wire together and nodes that are wired together tend to fire together. We discuss implications of this interpretation.

- **Issue #4: STAGES “[inappropriately] covers child development [and addresses adult pathology]”.**

We argue that, though STAGES by itself is insufficient as a full theory of child development, that its principles are validly extended for understanding child development. Also, all adult developmental theories explain earlier developmental levels using examples from childhood. We also discuss the validity of using STAGES concepts in discussions and practices about of psychological shadow and normal forms of adult neurosis.

- **Issue #5: STAGES “uses untested developmental stages.”**

As it is well known that O'Fallon has been “testing” STAGES empirically, we believe that what is meant is that STAGES is not “validated.” We discuss validity below. We correct errors in the description of the subjects used in O'Fallon's research. We also discuss how sentence length and complexity appear in later stages, and explain how STAGES relates to the “transcend and include” directive of development.

- **Issue #6: STAGES “presents problematic stage descriptions.”**

The Critique seems to be assuming that the Cook-Greuter descriptions cannot be extended or added to. But this is precisely what STAGES does, and part of its claim to having some added value. We respond to specific examples about the Expert stage, and discuss the use of “being and becoming” in STAGES.

- **Issue #7: STAGES “proposes a single metric to ‘measure’ orienting generalizations.”**

O'Fallon does not make this claim of the STAGES model. We elaborate in the article.

- **Issues #8, 9, 10, and 11: Related to Scientific Validity.**

Here we summarize O'Fallon's research study and results. Because of the limited availability of data and differences in the scoring methods for the highest stages (5.0-6.5), O'Fallon used different statistical methods for higher stages vs. all other stages. The results show (1) a strong replication with the scoring system used by Cook-Greuter up to Strategist (4.5), and (2) strong inter-rater reliability for the higher stages.

STAGES Research Overview. At the end of this article we give a summary of results from the yet-to-be-published full report on the STAGES validity study.

General Remarks. We also give general remarks on the Critique as a whole under four headings:

- **On Conjecture and Claims.** We discuss how the Critique, focusing on issues of rigor, in some ways lacks rigor itself, which may be acceptable for a “letter to the editor,” but is less so given the seriousness and potential impact of its claims.
- **On Commercialization and Transparency.** In keeping with our goal for a holistic single/double/triple-loop response, we discuss here an important issue not mentioned in the Critique: that Cook-Greuter and Sharma run a business that competes with O'Fallon's business. We acknowledge the tensions that arise in doing business and scientific reporting in parallel.
- **On Peer Review.** We acknowledge the importance of peer-reviewed publications, but also note that the authors of the Critique

have only a small handful of peer-reviewed journal articles, and, as far as we can tell, none that report on rigorous empirical research, as does the research to be published by O'Fallon.

- **Additional Remarks.** Finally, we comment on: (1) the fact that the Critique is framed as a general call for rigor in the field, but is aimed bluntly at only one theory; and (2) how the authors of the Critique distance themselves from it by withholding their names until its end. In the conclusion we step back and ponder self-reflectively on “how the hell did we get to this situation?”—in which our valued colleagues thought it necessary to draft a sweeping public disapproval. Terri and others working with her accept that they must share responsibility, and that the situation offers an opportunity for frank self-reflection. We offer our hope that this exchange will further theory, practice, and relational connections in the community.
- **Appendix 1.** Appendix 1 contains a response to the Critique by O'Fallon's statisticians, discussing some of the methodological fine points of the research.

In an accompanying article in this ILR issue, “On sentence completion assessments for ego development, meaning-making, and wisdom maturity, including STAGES,” Murray gives a summary of the STAGES model to aid the reader. That article also includes an overview of past research supporting the sentence completion test.

Preface and History

One thing we greatly appreciate about the community of theory and practice associated with Integral Theory is the invitation to be more fully human and more fully ourselves in our professional and scholarly interactions. We are blessed to be members of a community that welcomes multiple perspectives, cross-disciplinary mind-body-spirit approaches, and “triple loop” I/we/it action inquiry. Though it is difficult, we can at least aspire to engage in complex dialogues that include nuance about “truths” and fuzzy categories, acknowledgment of the possibility of our own biases and shadows, authentic vulnerability in sharing how the work lives in our bodies and emotions, and the sophistication of meta-speaking about the nature of a dialogue even as we are having it. In situations where there are contrasting viewpoints and complicated histories, these lofty possibilities may be particularly difficult to manifest, but at least we can relax in the knowledge that we are in a community that understands and tries its best to embody these many-facets of being human.

The critique of Terri O'Fallon's STAGES model, research, and dissemination written by Susanne Cook-Greuter, Ken Wilber, and Beena Sharma, gives us a rich opportunity to practice these things. The editors at ILR invited Terri to submit a Response, and Terri reached out to several colleagues to create this collaboratively written reply. The co-authors have all been deeply involved in STAGES theory or practice for some years.

Many readers will be familiar with the outlines of the historical background of this dispute, but we will summarize them. Susanne's work on Post-autonomous ego development extended Loevinger's work in very important ways, and stands as a major contribution to the larger body of work within the Integral Theory community and adult development scholarship. As her website says, her “twenty-five year research led to a reformulation of ego development theory and the expansion of the SCT to include stages of self-actualization and ego transcendence.”^[1]

Terri O'Fallon studied under Cook-Greuter's system and worked as a trained scorer for the MAP sentence completion test (SCT) for 6 years. Terri had previously co-founded Pacific Integral, which originally used Susanne's MAP method of SCT assessment. Terri then developed a derivative model called STAGES that added some levels, proposed a deeper explanatory model for how developmental stages emerge, and included a new method of scoring the SCT (based on the distinctions made in the new model).

Terri is also a long-standing contributor within the Integral Theory Community, and is deeply grateful for all she has learned from both Ken Wilber and Susanne Cook-Greuter, and for the significant support they have been to her in many ways over the years. In addition to her work in Integral Theory and Ego Development theory, O'Fallon's knowledge of human development has been informed by a lifetime working in a wide variety of teaching, counseling, administrative, and service professions. Based in her deep personal understanding of Integral Theory and Cook-Greuter's model, and her own experiences in scoring and working with people over a lifetime—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) might be explained at a deeper level using the ontological dimensions (or primordial polarities) that define the AQAL model.

Initially, Terri saw this intuition as a possible path to support both Wilber's and Cook-Greuter's work. A sense of service to these mentors, whom she felt deeply aligned with, fueled her to pursue the initial insight. She thought that an empirically validated use of the AQAL framework would contribute to a greater perception of its validity and soundness, as AQAL had come under critique and even attack from various authors. She also thought that it would be a beautiful tribute to Cook-Greuter's lineage if it could be shown that the ego-development model was indeed an integrally structured developmental model at its core.

From here the history becomes more multi-perspectival—that is to say, there may be different opinions on it, but we must briefly provide our version to set the context for our Response.

To ensure she was on a valid research trajectory, Terri had numerous (audio-taped) in-person meetings with Ken over several years, and many email exchanges with him (and has reviewed these documented encounters many times as she worked on her model and publications). Ken was very gracious with his time and coaching. Those discussions were crucial in fine-tuning many of Terri's initial intuitions and hypotheses. Terri also started working with two well-credentialed statisticians who helped design and evaluate her empirical studies. Terri checked in with Ken and Susanne on the design of the rather elaborate and lengthy research plan, and was not aware of any issues they had with either the STAGES model or the study design until after the study was completed and she started to privately share the preliminary results. She took these concerns to heart and thought that she had addressed Ken's main concerns along the way. She was surprised to learn, in seeing the draft of the ILR Critique letter, that Ken still had significant concerns.

For much of the period of developing the STAGES model and designing a research study to test it, Susanne and Terri were in communication and, like Ken, Susanne was very helpful. Susanne's concerns became more critical after the research activity was completed, and as early results were reported at the Integral Theory Conference. At the same time, as Terri published more about

the emerging STAGES model, and as STAGES grew in popularity, Susanne became increasingly concerned with some of the claims being made, on both theoretical and empirical grounds. Susanne and Terri engaged in many conversations (face to face and written) in which they attempted to come to a mutual understanding about these concerns. But in the end they have not resolved their differences—Susanne seems to feel that Terri has not addressed her concerns and Terri feels that Susanne has not understood her model and her detailed responses. This has been painful for both of them because of their long and close friendship, which they both make efforts to maintain in the midst of these differences.

In recent years it appears as though Susanne and her business partner Beena Sharma have been sharing their increasingly strong critiques and concerns about STAGES in private communications with many in the community. In a culmination of this trend, Terri was recently notified by the editors at ILR that Susanne submitted a detailed critique for publication. This lengthy Critique, which we have heard from several sources was in the making through private conversations (without Terri's knowledge) for about a year, was to be published within one week. Terri asked the editors to postpone publication for one cycle to give time to include a written Response in the same issue, which was generously granted by the ILR editor.

Apparently both Susanne and Ken felt that the existing modes of communication were not productive enough on key issues, and they took a more radical approach. We believe that all parties can share some blame for the imperfections in the communication process to date, and Terri, for one, regrets any un-intended ways she has contributed to breakdowns in the process, and assumes that there is something to learn in all of this that will help her improve how she communicates her ideas in the future.

The public publication of such a strong and direct critique is unusual within the established modes of communication in our community (notwithstanding the critiques of Integral Theory offered from outside the community). And it is especially rare among colleagues who have worked so closely and collaboratively for many years (and still offer kind words to each other even as this difficult chapter of their relationships unfold). We need and want to take this Critique very seriously and respond with due respect and diligence. Though both Susanne and Ken have voiced concerns and critiques over the years, the draft Critique recently received was the first and only clear summary of their issues presented in a form that allows for a direct and full response.

After the ILR editors showed Terri the draft Critique, Terri offered to Susanne and Ken to continue the ongoing dialogue and take another opportunity to clarify the misconceptions it contained, but she was put in a difficult position. She had already tried in one way or another to address these concerns at different times over the prior three to four years, yet misconceptions and differences remained. Now the language game had changed from collegial dialogue to public academic written high-stakes debate. To try again, at this new juncture, to offer corrections might only serve to broadcast her strategy for a Response, and by doing so, allow the final draft of the Critique to be strengthened, and her position, in this new competitive context, to be weakened. She offered to Susanne and Ken the option of (a) abandoning the published Critique and using the draft as a new starting point for dialogue, or (b) going forward with the publication of Critique and a Response. They chose the latter. That brings us up to date.

A summary of the STAGES model is given in a later section of the Companion Article in this ILR issue "On sentence completion assessments for ego development, meaning-making, and wisdom maturity, including STAGES" (Murray, 2017), to allow readers to understand the basic concepts in the model. The Critique lists eleven issues and describes some general themes, which we will address, starting with those that we agree with in whole or part. We do very much appreciate the generous statements given in the Critique's introductory section in the paragraph beginning with "There is much to appreciate about the STAGES model." We hope that the debate recorded in this ILR issue will serve the readership, create more mutual understanding among the authors involved, and move the field forward.

Before diving in to our responses, we want to acknowledge the significant contributions that Ken and Susanne have made—made to each of us authors of this article; made to all those familiar with their works; and made to those everywhere who have hopes that higher and deeper forms of consciousness will emerge within humanity on a large enough scale to take us through and beyond our significant challenges. It is impossible to acknowledge with enough breadth, or to bow deeply enough, to describe or honor Wilber's contributions, and suffice it to say that for each of us, our lives today would be very different and much impoverished had we not discovered his work.

For Susanne, we want to speak with gratitude to her ongoing role as a champion for ethical practices and academic humility within the community of integral theory and practice. For example, in Cook-Greuter (2013) she questions the tendencies of integralists and "evolutionaries" toward hubris, dangerous abstraction, unreflective privileged perspectives, confusing our elegant maps for the territory (despite our trying not to), and simplistic teleological assumptions about the "upward and onward" evolution of culture and community. As a respected elder scholar in the community, her stance on these matters has been of great value. We know that her main motivation in spearheading the Critique comes from a deep commitment to ethics and integrity, as it applies to both practice and to empirical research.

Acknowledgement of Issue #10

Issue #10: "STAGES misrepresents its validity: The STAGES theory and measurement are being prematurely promoted as validated"

First, we would like to acknowledge what we think is the most valid critique, which is the second part of issue #10 (we do not agree that that validity was misrepresented, only that it was promoted prematurely). Indeed, Terri was making verbal and written statements about the statistical validity of her model without having a research report available that would allow colleagues to scrutinize those claims. Terri has been working with two senior statisticians for five years, first to design and carry out, and then evaluate and write up, a series of studies to validate the STAGES model. In addition to testing its psychometric strength, the studies aimed to show that the STAGES method of scoring the SCT would result in scores that were essentially the same as those produced with the MAP scoring method—i.e. to show that the new scoring method replicated the scoring method it grew out of (as it added new features). The goal was to produce one or more papers in peer-reviewed journals. The process has been unexpectedly long, for many reasons.^[2] Once Terri learned from her statisticians that the results were positive, the results were mentioned in a peer reviewed ITC paper for which she was awarded an honorable mention. She then, enthusiastically began to mention these findings in talks and on the Pacific Integral web site.

The process of finalizing and writing up the studies took much longer however, and even as in-depth drafts that would support the claims were authored, Terri was constrained by the conventional rules of journal submission—journals expect new and unpublished material—so she was reluctant to put any of the specific results into the public domain, which might damage her opportunities to publish. (Though Terri did share much of this material confidentially with Susanne many months ago.) In hindsight, we can see how the premature announcements of final results was so problematic for those, like Susanne, who were skeptical from the start that Terri's method and model could be validated, and those, including Thomas Jordan and Sean Esbjörn-Hargens, who are concerned in general about the scientific rigor and integrity within the wider space of integral and developmental theories. The critique that such results should not be shared without publicly available documentation is completely valid.

And yet, the Critique offered is not completely valid even on this Issue (#10). After Susanne communicated her concerns about the claims on the PI website to Terri, Terri and the staff at Pacific Integral agreed that it was an issue. Well over one year ago, they then made every effort to remove language about the model or method being “scientifically validated” from their web site. They informed Susanne that these changes were made. The quotes offered in Issue #10 of Critique have been long revised and are no longer relevant.

We also wish to specifically mention the quote from Sean Esbjörn-Hargens, a supportive colleague of Terri's and a prominent voice in our community. He comments that “With ground-breaking work (such as with STAGES), I believe there comes a higher burden of responsibility...” We take this challenge to heart and appreciate his opinion on the significance (or potential significance) of this emerging model. His concerns are rightly linked to the issue mentioned above, i.e. that claims of research results were publicized without a research paper publicly available. We are not sure how much of his opinion is informed by a reading of the Critique itself, which, as we show below, contains misconceptions about STAGES, including an exaggeration of claims allegedly made by O'Fallon. One such misconception surrounds the requirements of sample sizes, and a justification for the robustness of the sample size is given later in this article (and in the Appendix). Though the publication of studies to date will clarify some of this, we agree that we also need to be even clearer on “which aspects have been fully or partially validated and which ones are still speculative,” and we discuss this in the Conclusion.

Another critique that we want to acknowledge as valid is that Terri erroneously claimed that Wilber supported or endorsed the STAGES model. Face to face conversations (many taped) and email conversations (on record) between Terri and Ken led her to believe this was the case.^[3] But the Critique publication makes it clear that Ken does not support as much of the model as Terri assumed. However, Terri does believe that some of the differences are due to misunderstandings of definitions of STAGES terminology rather than actual differences of opinion, and hopes that this article will help clear much of that up. As noted below, Terri made several changes to the STAGES terminology along the way, based on Ken's suggestions.

Responses to other Issues

Next, we will address the Critique's Issues in turn. We will note that, though the Critique is framed as a call to scholarly rigor, that there are few citations of STAGES material given in the Critique to ground its claims. This is understandable for an article sent as a “letter to the editor” rather than a peer reviewed paper, but the Critique would have been stronger with more citations. Much of our Response will correct misconceptions about the STAGES research and model that are found in the Critique. We acknowledge that there may in fact be some quotes from some of Terri's papers that could be interpreted, either through misstatement on her part, or misunderstanding on the part of the reader, to support the critique given. In the absence specific references, however, we are not aware of any that do.

Issue #1. STAGES “asserts specific descriptors based on a meta-model”

The Critique says: “Scholars agree that, based on a meta-model, one cannot make detailed assertions about specific stages on the developmental trajectory. Therefore, a core assumption of STAGES theory that specific ego stages belong to different AQAL quadrants and that ego development progresses in a cyclical and sequential pattern around the quadrants cannot be upheld.”

It would be helpful to see references to which the authors are making this statement. Nonetheless, we agree that, indeed, AQAL, and most of Wilber's significant theoretical contribution, is a meta-theory, a way of coordinating and integrating other theories (perhaps *all* theories about *any* aspect of reality). As claimed in Issue #1, it does not, and cannot be used to, make claims in specific meso or micro level domains. STAGES is not a direct application of AQAL, or an extension of its predictions. Rather, its structure was *inspired* by AQAL, and O'Fallon used the patterns contained within AQAL in a novel way (a way not predicted by AQAL) to develop definitions of perspectives at every developmental level. Her original intuition that the AQAL structure could be repurposed as a structure that repeats through developmental tiers was a hypothesis that research later confirmed as valid (at least to a first approximation). STAGES, unlike AQAL, is not a theory of everything, it is a model or theory of human meaning-making and perspective taking. AQAL describes human development, but is not a specific theory about the *underlying mechanisms* that drive the development of meaning-making and perspective taking, as is STAGES.^[4]

The Critique says, “according to AQAL theory, quadrants/zones don't ‘determine’ structures.” It is true that AQAL theory does not predict that its key dimensions of reality lead to the emergence of developmental structures. But it does not *prohibit* this possibility either. O'Fallon's empirical work seems to show that the AQAL dimensions can be used as a valid explanatory deep model for development.

O'Fallon believes that, as STAGES is shown to be empirically valid, it is a kind of empirical support for the power of AQAL primordial dimensions (or perspectives), while not being a direct support or proof of the validity of the AQAL model. STAGES 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. Wilber may disagree with the supposition that STAGES research is supportive of AQAL. Arguments for or against the validity of meta-models (like AQAL) should be based on the “data” of existing whole theories or models. In this we agree completely with Cook-Greuter and Wilber: STAGES research does not validate the AQAL model (it is supportive of it), it only validates the STAGES model.

Finally, the Critique says that “[as AQAL] and third-party research show, individuals at any stage may privilege any of the quadrants

as a base from which to observe reality.” This principle is refined, rather than refuted, in STAGES. It may be reasonable to assume that normally developed adults can look at objects “located” primarily *in* any quadrant/zone; and that they can look at any object (holon) that they are aware of *from* any of the 8 zones or methodologies (as noted in Wilber’s Methodological Pluralism (Wilber, 2006)). But in addressing this critique it is important to differentiate implicit *being/enacting* from *explicitly* making meaning of reality through language. A person acts within and through all four quadrants, since we exist as individuals in collective contexts, and individually and collectively we have interiors and exteriors—this is the “looking as” or inside zones of the four quadrants. The subject-to-object move of *explicitly* (through language) making meaning, i.e. “looking at,” conceptualizing, and thinking *about*, is a developmental achievement. (None of this is new to the Critique’s authors, but we are building up our case.)

First, the outside (objective stance) zones are not universally available in the same sense as the inside (subjective stance) zones. Second, we can note that individuals capable of only first person perspectives can’t take second person perspectives, and thus can’t explicitly take collective perspectives (and ego development and meaning-making development are primarily about what people explicitly think about and can verbalize). Third, the tiered structure of STAGES allows us to articulate that, though a person may look at or through any of the quadrants or zones, some people are constrained to make meaning of only concrete objects, and others of only concrete and subtle objects (and similarly for Metaware). For example, STAGES says that one can “see” *concrete* collectives from a second person perspective on up, but can’t “see” *subtle* collectives (e.g. cultural memes, as objects) until taking a fourth person perspective. One of the nuances made available through STAGES is that those at third person perspectives (Expert and Achiever levels), although they are in the *subtle* tier, are still seeing *collectives* from a *concrete* orientation. Thus, though all of the quadrant/zone perspectives are available to most adults, the types of objects (Concrete, Subtle, or MetAware) available to awareness are constrained by developmental level.[5]

Issue #2: STAGES “uses problematic terms to represent concepts”

The tiers in the STAGES model were originally called Gross, Subtle, and Causal, reflecting the stages of consciousness referred to in Wilber’s work and in a number of Eastern mystical traditions. O’Fallon had specific reasons for using these terms, based on her understanding of them. The description of the STAGES tiers maps well onto *certain aspects* of the Gross/Subtle/Causal model. But these words, especially Causal, have a variety of meanings in the literature.

Well over a year ago Wilber expressed his concerns to Terri about using these terms in the STAGES model, indicating that he held a different understanding of them. After a long discussion, Terri agreed that the terms may have been confusing or ambiguous, and changed her model accordingly. She now uses the terms Concrete, Subtle, and MetAware as the names of her three tiers. This change was made about one year ago, and has been used in the STAGES workshops, website, and other material. The change was communicated to Susanne and Ken many months ago. Thus, we think Issue #2 was outdated well before the authors chose to publish their critique, and we can only guess that there was a communication breakdown somewhere along the way on this issue.

Issue #3: STAGES “conflates stages and states”

The Critique claims that the STAGES model: (a) “continues to assert that stages and states correlate,” that it (b) “collapses [state and stages] into one, equating them,” and that STAGES (c) “demonstrates a serious misunderstanding and misrepresentation of Integral Theory.” The STAGES model does not make the errors in (a) or (b), and no quotes were given to indicate where these misconceptions came from. It is certainly possible that O’Fallon’s writings include elements that are open to such misconceptions, and could be improved upon. As to item (c), we believe that it is more the case that Cook-Greuter and Wilber misunderstand and misrepresent STAGES, as we give evidence of throughout this article. As we say elsewhere, STAGES research does not lead to claims about AQAL.

Critique issue #3 makes several statements, such as “one cannot ‘peak-experience’ structures,” almost all of which we agree with, which adds to the evidence that its authors misunderstand the STAGES theory. Also, the critique mentions things like O’Fallon’s use of the phrase “waking up is growing up... until recently,” that the authors acknowledge have changed. The “until recently” points to the fact that O’Fallon has modified her use of this phrase, exactly in response to feedback from Wilber. Many months ago, Terri informed others that these changes were made. So again, we wonder why past but corrected statements are included in the Critique, and are curious about the causes of the apparent communications failure.

On item (c), O’Fallon does seem to differ or modify Wilber’s notions with respect to the relationship between stages and stages. To have a somewhat different (but not diametrically opposed) opinion on the relationship between states and stages is part of a healthy scholarly dialectic—the type that can move the field forward. In differing O’Fallon does not “misrepresent” nor necessarily “misinterpret” Integral Theory, but simply has a different perspective on these constructs.[6]

Wilber’s “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 any state comes through one’s stage of development (Wilber, 2005). O’Fallon completely agrees. In addition, O’Fallon believes that experiencing certain states of mind are rough (perhaps not strict), prerequisites for the achievement of certain stable stages of development. The common interpretation of the Wilber/Combs principle does not seem to disallow this hypothesis. Neither the Wilber/Combs principle, nor O’Fallon’s corollary to it, has been proven empirically, and O’Fallon sees her belief a hypothesis.

Wilber’s use of states draws largely from Eastern mystical and contemplative traditions. However, a “state” of mind (or mind/body/spirit) is a much broader concept than is captured by the major states of “Waking, Dreaming, Deep Sleep, Witness,” etc. (or by Gross, Subtle, Causal, Turiya, etc.), spoken of in Eastern traditions. Also, for integralists, the term *state* is often used to refer to some sort of special, altered, or non-ordinary state that, furthermore, has spiritual implications. We prefer to use “state” in a more general sense. Attention, focus, awareness, disorientation, openness, flow, timelessness, bliss, these are all states of mind or consciousness as well, and may factor into the mechanisms underlying human development and meaning-making.

In the Companion Article we offer a definition of states and stages that we think is compatible with, but different from, Wilber’s description of the Wilber-Combs Matrix—one that is also congruent with modern cognitive and brain theories. “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*. The upshot of 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) (Siegel, 2012). This leads directly to O’Fallon’s conjecture that repeated experience of certain states may be necessary (but not sufficient) for attaining certain stages. States that become stable experiences (sometimes called “stage states”) are one ingredient of stage attainment. It is a straight following from common cognitive neuroscience theory, given her more expanded and non-spiritualized meaning of “state.” (See an extended version of this argument in the companion paper.)

A few additional points can be made on this subject. As early as 2006, Wilber seems to imply that, in some sense, states precede stages: “Once you stably reach a stage of growth and development, you can access the capacities of that stage—such as greater consciousness, more embracing love, higher ethical callings, greater intelligence and awareness— virtually any time you want. Passing states have been converted to permanent traits” (Integral Spirituality, p. 5). Wilber is also quoted in Darrall-Rew & DiPerna’s book, *Earth as Eden* (2016, pp. 199, 120), as saying that some stages may require certain states.[7]

O’Fallon has been writing about this relationships between states and stages since 2013, and has begun using “vantage point” language from Dan Brown and Dustin DiPerna’s work, to refer to the major states (state stages) that have become ordinary and are necessary but not sufficient for particular stages to arise. She recently noticed that in Ken’s new book he is referencing vantage point terminology, and equating his “Gross, Subtle, Causal, Turiya, Non-dual” language with the Vantage points, indicating more areas of potential agreement.

Finally, it is important to note that states are not inculcated into the primary factors in the STAGES scoring system (as are the three primordial dimensions of the AQAL quadrants/zones). Thus, this state-vs-stage issue has less bearing on the actual scoring system or on the validity of the STAGES model research.

Issue #4: STAGES A.”[inappropriately] covers child development within adult development theory,” and B. it inappropriately addresses adult pathology

The study of child development is, in many ways distinct from the study of adult development—at least in academic scholarship. There are different concerns involved as the brain matures biologically, especially in early childhood development. Yet, as far as we can tell, every adult developmental theorist makes analogies to childhood behaviors and world-views in describing the earliest stages of development. These are actual analogies, not metaphors, as adults displaying the lowest levels of development (whether habitually or in regressed states) exhibit developmental levels similar to those seen habitually in children. A first-person perspective is a first-person perspective, regardless of the age of the actor.

In her white paper, *Nine Levels of Increasing Embrace* (2013), Cook-Greuter describes the early stages (e.g. Impulsive) in terms of how children behave. In addition, Loevinger’s colleagues applied ego development to pre-teenagers. The neo-Piagetian adult developmental research of Fischer and Commons can be directly extended to any age (and even to animal cognition). Given all of this, we see no reason that O’Fallon should not mention child developmental phenomena in her explanation of STAGES.

Also, there is nothing to prevent the STAGES model (or Cook-Greuter’s or Loevinger’s models) from being used to *measure* the development of children. In fact, O’Fallon hopes to apply STAGES to this domain. However, we acknowledge that there may be special considerations for applying adult developmental theories, and especially sentence completion tests, to children. For children, these models and methods may be less useful, or not applicable for many important questions. For example, linguistic competence is needed for taking the sentence completion test. However, since STAGES is a theory about underlying perspectival drivers of development, its application is not tied to the sentence completion test. Loevinger’s assessment is tightly tied to the examples given in her scoring manual, and Cook-Greuter’s scoring method releases that constraint somewhat, but is still example-based. The movement from exemplar-based to model or principle-based scoring is one of the important advances claimed for STAGES.

Issue #4 includes two separate issues. The first questions the application of STAGES to children mentioned above, and the second questions the application of STAGES to psychopathology and clinical psychotherapy. The authors note that “Loevinger... explored healthy adult development as a response to the prior over-emphasis on psychopathology. Her WUSCT... is normed only on reasonably functioning adults,” and suggests that STAGES application is over-extended into such areas. We believe that this critique is pointing to the material presented in STAGES workshops offered by O’Fallon and Barta. Kim Barta is a holistically-oriented integrally-informed psychotherapist, and, largely through his influence, these popular workshops weave concepts of psychological health and shadow work deeply into the description of the STAGES model.

Largely through Wilber’s influence, the discussion of adult development within our community has always been about both development and shadow work (growing up, waking up, and cleaning up). The workshops offered by Cook-Greuter and Sharma discuss shadow as well. Though Barta’s expertise allows him to introduce concepts from psychology and psychotherapy, the aim of the work is to help “reasonably functioning adults,” and those coaches and therapists working with them, to understand and navigate the territory of the typical psychological issues and common neuroses of adult life.

O’Fallon and Barta make no claims that the psychological principles and models that they use, nor their prescriptive suggestions related to personal healing and clinical work, are scientifically validated. Those working in leadership and organizational consulting use a variety of models that have not been scientifically validated but that offer useful meaning-making frameworks and tools for reflection (for example the Polarity Management framework used in Cook-Greuter and Sharma’s workshops). As in most general workshops that might be attended by coaches, therapists, or other clinicians, professionals must use their own judgment and professional training to discern how to apply the material from the workshop to their work.

The empirical validation of STAGES is in the realm of assessment using the SCT, not in the realm of models shared in these workshops. Models offered in workshops are, like Wilber’s theory in general, meaning-making frameworks whose validity is best measured by the “functional fit” criteria Wilber suggests for the lower-left quadrant—i.e. if they provide cognitive frameworks that help people heal and grow themselves and others (i.e. waking up, growing up, cleaning up, showing up, etc.), and they cause negligible

harm, then they are “good,” even if not scientifically validated.

Issue #5: STAGES “uses untested developmental stages” (and RE transcend and include)

We are not sure what this means, as the research design, as seen early on by Wilber and Cook-Greuter, was meant explicitly to “test” these additional stages.[8] Prior to the results of the empirical study, STAGES did indeed contain levels that were untested. But, the question of whether the study produced *valid* results aside, the authors of the Critique knew about the study to test these levels—so these stages were certainly “tested.” Perhaps what the authors meant to say was “STAGES uses un-validated developmental stages.” We address validity at Issue #8.

In addition, the Critique says that “to get responses from the new stages, O’Fallon sought out individuals who, in her estimation, could produce such responses—often second-time test-takers and spiritual exemplars”. First, the authors do not cite any evidence for this claim. The identity of the individuals used in the study has been kept strictly confidential according to both the study IRB and common ethical principles. Second, similar to the statistical method of “stratified sampling,” it is valid to seek out data that covers a full range of the phenomena being studied. O’Fallon wanted more high-scoring individuals and individuals likely to score higher were recruited (using a web-site method that was recommended by Cook-Greuter). Because the study does not make any claims about how representative the study population is of the general population, this targeted recruitment has no bearing on its validity. We acknowledge that the pool of advanced-stage individuals who asked for or made themselves available for STAGES assessment may not be fully representational of the general population of advanced-stage individuals. This is a challenge for most developmental research, a field which includes extremely few random population sampling methods, and future studies at larger scales may improve this limitation. In addition, it is not uncommon for research into contemplative practices or spiritual mentoring to enlist experienced practitioners.

The critique also says: “The dozen or so sentence completions she shared with Cook-Greuter were all hyper-complex, and many mimicked insights and phrases from the spiritual literature. The STAGES results showed no recognition of the possible simplicity-after-complexity that can come with increased maturity.” The STAGES model claims that with each person-perspective, a new type of object is seen and understood, and that each person-perspective has first a passive stage and then an active stage. Cook-Greuter’s understanding of late stage development seems to be that, starting with Construct Aware (i.e. in what O’Fallon calls the MetAware tier), sentence completions become by and large more simple or elegant. In the STAGES model, this in part relates to the passive (first) part of the first person-perspective. Completions in the active MetAware levels (5.5, 6.5), however, can indeed be more complex (though not necessarily). Within the stages data the length of completions continues to increase through 5.5, and then drops off starting at 6.0.

O’Fallon was originally trained as a scorer in Cook-Greuter’s model, and, as a member of that scoring community, has seen its internal workings. Terri had noticed a category of sentence completions that had an uneasy relationship to Susanne’s scoring method. These were complex sentences that seemed to have Construct Aware (5.0 in STAGES) elements, but were not ultimately counted as Construct Aware because they were not simple and elegant enough. For Terri, one of the indications that the STAGES model was useful was that it explained and dealt with these completions by placing them comfortably in her new 5.5 category.

O’Fallon had a conversation with Cook-Greuter in which she was trying to explain the difference between passive 5.0 and active 5.5, and probably did focus on the more complex examples because, those scored at below 5.0 in Cook-Greuter’s model and above 5.0 in STAGES. What determines a 5th person perspective in STAGES is whether the objects are MetAware objects (e.g. awareness of awareness, or very subtle objects), regardless of the complexity of the sentence.

Issue #5 includes a second critique: “the available STAGES data does not allow one to draw general conclusions about the ‘must include and transcend’ nature of subsequent stages—a basic tenet in the field of developmental theory.” All adult developmental theories agree or assume that each stage includes and transcends those prior (though the terms used vary). Neither Loevinger’s nor Cook-Greuter’s work makes any attempt to “prove” or validate this assumption. Empirical work in the field has shown conclusively that they follow each other sequentially (for those levels where enough data has been gathered), but the “transcend and include” structure is basically an assumption borrowed from Integral Theory and “validated” through the force of explanatory reasoning, not empirically.

On the other hand, the theories of Commons (Hierarchical Complexity theory: Commons & Richards, 1984; Commons & Pekker, 2008) and Fischer (Skill Theory: Fischer, 1980; Mascolo & Fischer, 2015) have quite explicit notions of transcending and including. For example, levels in Hierarchical Complexity Theory are non-repeating recursions that coordinate, organize, or transform lower level elements. In these theories, this aspect is embedded into the very definition of development and what it means to be a stage. They are not added on for explanatory purposes. Kegan’s developmental theory (1980) is also explicitly a transcend-and-include theory, and the subject-to-object sequence of increasingly complex stages each build explicitly (by definition) upon those prior.

The STAGES model is closer to the neo-Piagetian Hierarchical Complexity (HCT) and Skill Theory (ST) in this regard. Like these other theories, STAGES is model-based, rather than the example-based underpinning of Loevinger’s work (see the Companion Article in this issue). In STAGES, the progression of levels is based on underlying principles that give rise to the sequence of levels. The progression from Concrete to Subtle to MetAware is mostly about the type of object one is aware of, though one could make a strong argument that this progression involves transcending and including, it is not a necessary component of the definition of the terms. However, the progression of passive>active>reciprocal>interpenetrating within each Tier *is*, by the very definition of the words, a transcend-and-include sequence (each building upon the prior to coordinate, organize, or transform).

Issue #6: STAGES “presents problematic stage descriptions”

It appears that what is meant by “problematic” here is that O’Fallon’s description of stages do not agree exactly with Cook-Greuter’s. The Critique authors may be assuming that the Cook-Greuter descriptions cannot be extended or added to. But this is precisely what STAGES does, and part of its claim to having some added value. Because it is model-based, the model does indeed predict stages and their characteristics. Are these predictions correct? As we discuss in the sections on validity, our study indicates that they are correct in that, at least up to 4.5, they correspond well with prior models (because the scoring system is validated, and the scoring

system is derived directly from the model). (In the Companion Article at the end of Section 6, we further discuss how the two models might be expected to measure slightly different phenomena.)

Issue #6 includes:

Experts (Stage 3.0 in STAGES theory) are said to have a capacity to step into other people's shoes (i.e., to take their perspective). This is not supported by the five decades of evidence from the Loevinger/Cook-Greuter's research. It seems that, according to STAGES, having feelings for others—which, as research shows, young children can definitely have—is conflated with being able to take on someone else's perspective. The ability to take another's perspective is a different and more adult capacity that integrates cognition and feelings. This ability is not yet formed at the 'Expert' mindset.

The authors do not cite a source for this, and it is a misconception about STAGES. O'Fallon believes that the STAGES interpretation of the Expert level is actually very close to what Cook-Greuter teaches. "Feelings for others" does not enter into the STAGES description here. Though Experts (3.0) tend to see things only from their own perspective, they have the capacity to think about what they would do if they were in the other's situation, and to try to take the other's perspective. The limitation is that when they try to do these things they can't see that they are still coming from their own perspective when they do this (they can't "see" or consider that their own perspective affects the process). The outcome is that they argue from a single perspective, their own, on both "sides."

Issue #6 includes:

For example, stages are assigned alternatively as stages of "being" or "becoming". This does not align with existing theory and observations. All stages can be seen as having components of "being" and "becoming."

We agree that there are aspects of being and becoming to all stages, in fact to all phenomena. The words "being" and "becoming" are sometimes used as part of the explanation of the Passive vs. Active stages in the model. So yes, it does not align with "existing theory," i.e. Cook-Greuter's model. It does align with "observation" to the extent that the STAGES model is a valid replication of prior models (as argued for later).

In the STAGES model, some stages background being and foreground becoming, and vice-versa, so one is more prominent than another at times, therefore, there is no real disagreement here. So, it does align with "existing theory and with "observation" to the extent that the STAGES model is a valid replication of prior models (as argued for later). I.E. the active (becoming) manifestations of a person-perspective do follow the passive (being) manifestations. Note also, for example, that 3.0 is a passive (being) orientation to subtle objects only, not to everything, and that 3.0 continues to maintain the active orientation to concrete objects established in the first Tier.

Issue #7: STAGES "proposes a single metric to 'measure' orienting generalizations"

STAGES theory does not make this claim (again, a specific reference for where they got this impression would help). By "orienting generalization" we assume the authors mean the AQAL categories, and especially the quadrant/zone categories that are used to describe both primordial dimensions of reality and methodological perspectives that one can take upon any object within reality.[9] Wilber's orienting generalizations are used as ontological categories in his "theory of everything," and as mentioned with Issue #1, STAGES is not a theory of everything but a model of the development of meaning making and perspective-taking (or ego-development—as are the models of Loevinger and Cook-Greuter). The basic STAGES assessment, which yields a "center of gravity" measurement, does not place people, their interiors, or performances "in" any of the quadrants or zones of the AQAL model, nor does it measure states or types—it yields only a "level", in AQAL terms (of course all such assessments are approximations and simplifications).[10] The AQAL orienting generalizations, as ontological categories, are pre-givens, and as such cannot be "measured" at all, but are merely assumed (or used provisionally).[11]

As is explained in the Companion Article, STAGES does use the quadrant/zone parameters for its underlying explanation of development sequencing. For example, STAGES says that the early 3.0 (early Expert) level deals with subtle or abstract objects from an *Individual, Inside, Exterior* perspective. But it does not try to classify the whole person or performance into the ontological orienting generalizations of Individual, Inside, or Exterior. Within the sentence completions one can notice preferences for seeing the world through certain of the quadrant perspectives, but again, this is an indication of a psychological (meaning-making or perspective taking) style, not an orienting generalization about the nature of reality or of any object/event within reality.

Cook-Greuter (1999) made a significant contribution to ego development theory when she recast Loevinger's levels in terms of a sequence of person-perspectives. This innovation was used to frame the entire model but it was not built into her scoring system, which was designed as a continuation of Loevinger's scoring manual. In a related innovation, Wilber (2006) describes his Integral Perspectivism:

...[the eight] primordial perspectives [of any occasion]—the inside and the outside view of a holon in any of the four quadrants... Each of these zones is not just a perspective but an action, an injunction, a concrete set of actions in a real-world zone [that] discloses the phenomenon that are apprehended through the various perspectives...'perspectives' simply locate the perceiving holon in AQAL space... (p. 34-35).

Wilber's model (including his Integral Math system) frames perspectives in terms of the quadrants (individual/collective and interior/exterior) and first-person (inside) and third-person (outside) perspectives, but does not speak directly to development (though his overall AQAL Theory does of course include development). Wilber's definition of "perspective" and Cook-Greuter's use of the term are compatible (and they reference each other's work strongly), yet different and not fully integrated. O'Fallon saw her STAGES model as creating an integration between these two models by describing Cook-Greuter's sequence of perspectival levels in terms of Wilber's ontological categories, and designing a scoring system based on the quadrant/zone categories (which was later validated empirically—see below).

Continuing the quote above from Wilber: "To take such-and-such a perspective is to be arising in this particular area of the AQAL matrix (In fact we will soon give the address of a holon in the AQAL matrix as address = altitude + perspective, where altitude means

degree of development and perspective means the perspective or quadrant it is in.)” STAGES builds upon, but does not directly follow, AQAL because, using Wilber’s published definition of “address = altitude + perspective,” it turns this around to formulate “perspective = address + altitude,” where the terms have a slightly different meaning in STAGES.

Though meaning-making is a wide “line” of development covering many sub-capacities, STAGES does not measure most of Gardener’s “multiple intelligences,” the physical world, or anything broader than human meaning-making. However, being a “wide” line, meaning-making and ego development, as constructs, do overlap with related constructs including spiritual intelligence, social/emotional intelligence, and reflective reasoning, and thus these theories do contain many “orienting generalizations” or sweeping meta-principles about human nature. (See the Companion Article for an in-depth analysis of the meaning of the construct of meaning-making.)

Finally, STAGES does not use “a sentence-completion test [to] test for states,” as the Critique claims. However, there are recognizable descriptions of vantage points in people’s sentence completions, and vantage points have state implications. STAGES scoring notes when someone clearly describes a state that has become ordinary (a vantage point). Vantage points are an aspect of one’s perspective taking, and as such don’t reveal the state that a person is in as much as a way of understanding the world that is made possible from practiced and stabilized state experiences. While STAGES scoring is based on four primary dimensions (Individual/Collective; Passive (Inside)/Active (Outside; Exterior/Interior; and Concrete/Subtle/MetAware), vantage points can be secondary stage indicators, indicative but not sufficient to determine a stage.^[12]

Issues #8, 9, 10, and 11: Related To Scientific Validity

Issues 8, 9, 10 and 11 are related to the experimental studies of validity conducted by O’Fallon and colleagues:

8. It uses an unproven measurement method.
9. It lacks scientific validity and confirmation.
10. It misrepresents its validity: The STAGES theory and measurement are being prematurely promoted as validated.
11. O’Fallon presents STAGES with a degree of certainty despite known issues and lack of scientific rigor and research.

Claims that STAGES is unproven, lacks validity, misrepresents its validity, is without rigor and research, and is presented over-confidently are pointing to the same questions, and we treat them here as a whole. We addressed Issue #10 at the start of this article, by noting that (a) claims to validity were indeed made after preliminary results were available and before research details were publicly available, (b) publication of results has taken much longer than expected, and (c) such claims were removed from the web site in agreement that claims should be supported by publications.

Details of preliminary results were shared with Cook-Greuter, either in person or in writing, many months ago. In her responses there were few specific questions or challenges about the details of the research design or the analysis methods. O’Fallon’s preliminary public claims that STAGES had scientific validation were indeed “unproven” for those who had no access to a detailed description of the research (see Issue #10 above), but the situation is different for Cook-Greuter and Wilber, who have had access to descriptions of the study. From them we would like to hear of specific problems with the research design and statistical analysis methods. Of course, wider critiques will be possible once the research report is published. In the meantime, O’Fallon can share confidential late drafts of the unpublished research results submission with colleagues who request it. We hope that subsequent drafts of the research paper will be strengthened by comments from others.

Based on the Critique and her conversations with Cook-Greuter, O’Fallon believes that Cook-Greuter maintains a misunderstanding of the scope of the STAGES study design, that it is merely an inter-rater study. Inter-rater studies are used to check the validity of a scoring method—to make sure it is sufficiently objective and reproducible. The STAGES research includes this, but is more broadly a *replication* study which tests whether two scoring methodologies produce equivalent results (or nearly so).

The Critique says

“To propose that a single interrater-reliability study confirms the STAGES hypothesis is unscientific” and “STAGES cites one small study (~150 subjects) that used a sentence completion test to show good interrater-reliability for its measure, compared with the established, manual-based method. Typically, a study would need at least 500-1,000 subjects to support the kinds of claims being made.”

In response: first, as mentioned, the research is not merely an inter-rater study, but a replicability study. Second, 150 subjects can in fact be substantial even for inter-rating (as well as for replicability). Many peer-reviewed published studies in related fields are based on far fewer individuals and can still produce strong statistical results, and a quick scan indicates that many of the Loevinger-tradition research studies use fewer than 150 subjects (see meta-analyses in Cohn & Westenberg, 2004; Manners & Durkin, K., 2001). Our statisticians in fact claim that 150 is quite a substantial sample size compared to similar research. In addition, as explained in the Appendix, 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. The Appendix includes a letter from O’Fallon’s statistician, with some details about the methods used and what a replication study is.^[13]

There also seems to be some misunderstanding of the statistical methods used. Those in the Loevinger tradition often use a correlation statistic (usually Pearson’s) to compare scores, but contemporary statistical theory shows that the Kappa statistic is more appropriate and demanding for comparing categorical items. In the Appendix Dr. Polissar discussed this difference at length. For a given study comparing categorical (or ordinal) values, a Kappa value will almost always be lower than an accuracy statistic, and what is considered a good Kappa statistic is lower than what is considered a good correlation statistic. Cook-Greuter may have misinterpreted the Kappa values found in the study as being less impressive than they actually are by conflating them with the less rigorous correlation statistic.

Cook-Greuter has claimed that the size of the study is too small to draw valid conclusions. One does need a large number of examples (data points) to construct an example-based scoring system such as hers and Loevinger’s, because of the vast diversity of

linguistic productions that are possible at each level. Given that ego development research and theory are traditionally derived directly from SCT scoring results, these also, in effect, require vast amounts of data. The same is not true for model-based theories like STAGES. A validation that STAGES scoring replicates Cook-Greuter/Loevinger scoring, or that it has valid psychometric properties, can use common statistical methods. The statistical significance of the results is a function of the amount of “noise” or random effects in the phenomena (and measurement). It is possible to generate highly confident results with a relatively small amount of data, assuming that the data is representative of the population. We have no reason to believe that the approximately 150 inventories randomly chosen for this study are not representative of the full set of protocols.

In the end, a rigorous critique of STAGES empirical research should involve a conversation between professional statisticians on both sides, as neither Terri nor Susanne (nor Ken) have the necessary background for this. We would welcome this level of peer scrutiny, now that the research report is near a final submission form.

Regarding the Critique’s “no research yet supports the STAGES claims for childhood development or for the proposed highest stages at the upper end of the scale.” First, as we note elsewhere, O’Fallon has not claimed, nor is there yet a research study, that any empirical conclusions about children can be made using the STAGES model (though a study with children is planned). Second, the research results do in fact show validity for upper stages (see below).

Next, the critique claims that the scorers (“co-observers”) in O’Fallon’s research are “her colleagues at Pacific Integral [and should] be considered biased... invested in supporting, confirming, and disseminating the model and PI’s approach.” The four scorers did indeed have close relationships with Pacific Integral. However,, if we assume that they did not bluntly cheat by talking to each other about their scores (which we of course do claim to be the case), then this should not bear on the results. In many research studies that use inter-raters the scorers are closely associated with the research group, and are often graduate students working closely with the group. Wanting a study to succeed (or being invested in the outcome) cannot make one a more accurate or more highly inter-rated scorer.[14]

It is certainly the case that a scoring method is considered more robust as more scorers are trained and certified as being accurate. In the three years since the STAGES research was done, there are now about a dozen certified or soon-to-be certified STAGES scorers. Again, we emphasize and agree that STAGES is a new model and we have much to learn, and it is always possible that future research will not show results as strong as the recent replicability study.

Regarding the (long removed) PI website claim that “STAGES is the first statistical validation of Wilber’s integral theory”. First, the question of whether STAGES itself is “validated” has been discussed. Second O’Fallon has changed her perspective in the last two years about whether what she is doing validates Wilber’s Integral Theory. She has diminished her initial zeal on this point, and now believes that her research only “supports” the validity of Integral Theory. A meta-theory of orienting generalizations cannot be validated through empirical study, but rather its validity is in its usefulness and meaning-generative capacity, and in this way the strong results from STAGES research and its continued applications support the validity of Integral Theory (and specifically the quadrant/zone part of AQAL).

As to item #11, we know that the phrase from O’Fallon (2013) describing STAGES as “an AQAL periodic table of consciousness, a prophetic structure that points to the developmental trajectory of an Integral human” was problematic for Cook-Greuter. A clear justification for the admittedly extravagant use of the “periodic table” metaphor is given in the Companion Article. The use of the periodic table as a metaphor for models in the human sciences is found in many places, and is understood to be metaphorical.[15]

STAGES Research Overview

STAGES has proven itself to be very useful at the level of metaphor and model, i.e. as a meaning-generative tool. As the Critique generously states: “STAGES has inspired and engaged many others to support research into human development at all ages, capabilities and proclivities.” But that does not imply that the STAGES scoring model is scientifically valid, which can only be shown through empirical studies. Below we will 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).

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 makes the systems difficult to compare using replicability methods. For these highest levels an inter-rater-reliability study was done. Thus the second main hypothesis was that STAGES scoring at these levels had adequate inter rater reliability (IRR).

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).

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 cross-check O'Fallon's CG/L-system scores. Second, each of the four scorers was compared separately with CG/L scores.[16]

The replicability of all four STAGES scorers (for levels ≤ 4.5) was found to be in the "excellent" range[17] 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 higher levels. Thus, the second hypothesis was confirmed. 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 containing sufficient detail to back up such a claim.

General Remarks

Below we make some general remarks, some referring to material in the Critique that is outside of the 11 "Issues;" and we also comment on the situation as a whole.

On Conjectures and Claims

The Critique includes the following:

- "[it] behooves all of us to be careful with our claims."
- "In scientific research, the originator of an idea holds the responsibility to distinguish between theory and conjecture, and to not make unwarranted claims,"
- "A practitioner's qualitative observations and anecdotal evidence, unsupported by empirical data, are not sufficient to claim validity for any scientific model or specific aspects of that model."
- "[one should] exercise great caution precisely because this area of research is so important and impactful," and
- "Instead, this critique points to what can go wrong when the peer-review process is lax and creative theorizing is accepted without scrutiny."

Though a "Letter to the editor" is not a scientific paper and is not expected to be a rigorously written as regular articles, we think we have shown that the Critique shows substantial "shooting from the hip," and includes some un-substantiated claims about STAGES, itself falling prey to some of the problems it attributes to STAGES reporting. We have granted that there was a period during which O'Fallon and colleagues made public claims too far before publication of results. However, the research design and analysis have been rigorous.

As to this claim:

"To offer the STAGES model as the most advanced in the field of human psycho-spiritual development. A model that supersedes (yet claims support from) Ken Wilber's AQAL theory is, to say the least, a vast exaggeration..."

This certainly is a vast exaggeration, a vast exaggeration of what O'Fallon says about STAGES. STAGES has positioned itself in the Pacific Integral marketing contexts as a definite advance over prior models and scoring methods. In the context of more objective reporting, we can say that it has *some* advantages, while having some drawbacks, including its newness (and therefore tentativeness) as compared to the MAP system used by Cook-Greuter. No sources were cited for this supposed exaggerated claim, so we are not sure why anyone would believe that Terri thinks her just-emerging model of human meaning-making development supersedes Ken Wilber's seminal meta-model of all of reality.

We have shown that much of the Critique is based on misconceptions about STAGES, with (as far as we can tell) only a small number of actual real differences of theoretical opinion. Like Wilber's AQAL and Cook-Greuter's ego development model, STAGES includes levels of detail and complexity that require some effort to assimilate. Even given the many conversations O'Fallon has had with the authors of the Critique, it is understandable that their knowledge of STAGES is partial, in part because only a few articles have been written describing it, and those do not cover it in the depth given in STAGES workshops (of which there are several levels). But it seems that its authors believed that they did understand STAGES well enough to publish their critique.

On Peer Review

The Critique says "We hope this memo marks the beginning of an ongoing effort to improve the peer review process as it pertains to theories from the Integral community." We agree with the Critique that the peer review process for substantiating scholarly claims is very important for the progress and ethics of any community of theory and practice. The Critique is framed in the context of peer review and its importance, yet it exists in an uneasy relationship to peer review. It is written as a Letter to the Editor and is not itself peer reviewed. We do take the point implied in the Critique that, without a research publication by O'Fallon, there is no article that can be peer reviewed (and one is soon to be submitted). In a sense, the Critique is a review of the entire STAGES *project*—by some scholarly *peers*, but is not a typical peer review of a publication or of a research study. It itself lacks the rigor required for that purpose.

Producing a peer-reviewed report of her empirical research study has been high on O'Fallon's priority list for several years. This has taken much longer than expected as we mentioned, and is now close to being submitted. We summarize the results of that study below.

O'Fallon does in fact have peer-reviewed articles within the integral scholarship community that describe theoretical aspects of

STAGES (and a number of other peer reviewed publications not focusing on STAGES). O'Fallon (2011) and O'Fallon (2013) were peer-reviewed papers presented at Integral Theory Conferences, and both were given awards recognizing O'Fallon's significant contribution.

We would like to note that the academic peer review process, while having important goals, has many documented flaws (see Ioannidis, 2005; Fleiss, 1981; Jefferson et al., 2006). We join with those who hope that integrative and second tier perspectives will help create more holistic and evolutionary versions of this process. O'Fallon struggles with some of the same issues that Wilber and Cook-Greuter do, all being "independent scholars" not working directly under the umbrella of an academic institution, and not benefiting from the structural or financial support for scholarship that comes with those positions (the tradeoff being that they are able to engage freely in research and applications that are outside traditional academic silos and constraints). In fact, in part due to these limitations for independent scholars (though the Loevinger lineage of ego-development research has an extremely robust academic footprint), both Wilber and Cook-Greuter have only a small handful of peer-reviewed journal articles, and, as far as we can tell, *none* that reports on rigorous empirical research.

On Commercialization and Transparency

In keeping with our goals for including multiple relevant perspectives, we will mention what was not mentioned in the Critique: that Terri and colleagues at Pacific Integral run a business that, in part, competes with The Center for Leadership Maturity, which is run by Susanne and Beena. We cannot know how much this fact contributed to their decision to publish a direct critique of STAGES, but we must assume it has some conscious or unconscious contribution. The business aspects of this situation are certainly a motivator for most of the co-authors of our Response to be putting substantial time into the effort. Fitch and Barta are business partners with O'Fallon, and Murray founded a company (Open Way Solutions) that features a technology for automated scoring of the SCT based on the STAGES model. Both as defenders of STAGES and as owners of businesses, we do not claim, implicitly or explicitly, to have an objective viewpoint on these matters. This undoubtedly introduces biases that we invite the larger community to comment on.

Because STAGES is used in O'Fallon's businesses, and STAGES research is thus far not carried out by independent organizations, we acknowledge that there will be tensions between the objective reporting of scientific "facts" and the desire to promote, publicize, and extol the value of the STAGES model. We ask the community to "keep us honest" in this regard with frank feedback, and we appreciate that Susanne did this many months ago and continues to do so (and this lead to past modifications to the PI web site as mentioned).

Returning to the topic of peer review, we can note that peer review ethical standards, such as the COPE standards (subscribed to by the Journal of Adult Development where Cook Greuter is listed as a reviewer), have a clear policy on declaring potential conflict of interest.^[18] The Critique skirts around such policies because it is not itself a peer review of another paper, and is presented in the more informal form of a Letter to the Editor. But we believe that in such a context transparency on matters of conflict of interest is still important.

The Critique says "Given the general pressures to gain market share in our field, overpromising scientific validity and premature dissemination are not unique to advocates of STAGES." Though we do not agree that STAGES outreach has "overpromised scientific validity," the needs of running a business, or even the needs of promoting one's ideas in the mimetic "marketplace" of scholarship, are often in tension with the need to be objective in scientific research and reporting. Indeed, these tensions are ubiquitous in scientific research (e.g., see Latour, 2005; Ioannidis, 2005). From an integral perspective we can move away from simplistic notions of "objectivity" and "proof" and address the practical nuances and tradeoffs, and see these tensions in terms of ongoing processes and "polarities to be managed" rather than "problems to be solved." This requires ongoing feedback from many types of stakeholders and ethical oversight from outside parties.

The STAGES model is being presented increasingly as a conceptual model with pragmatic principles that can guide consulting, therapy, personal growth and other applications (just as Erikson's model, Cook-Greuter's MAP, and many other psychological models have been used). As in every model that we know of that is used in these contexts, much of what is proposed or recommended in workshops is not scientifically validated, and is not meant to be. Given the results of the published study, when it comes out, it seems fair, particularly in workshop and business contexts, to say that STAGES is empirically or scientifically validated. However, for those with inquisitive, probing, or scientific bents, this begs the question of what "validated" means. Certain aspects of it have been shown to be valid to a certain degree, so far. No theory can ever be "proven" for all time, and STAGES in particular is a new theory with a very short history of empirical validation. In fact, much of its validation rests on the hundreds of studies showing the validity of the sentence completion instrument for assessing ego development coming out of Loevinger's seminal work (see the Companion Article). O'Fallon is committed to a dynamic evolutionary process for theory/model development, as is articulated in "grounded theory" theory (Haig, 1995; Hussein et al., 2014; which shares principles with the "Action Research" of Torbert, 2011). The STAGES model is expected to change over time based on new evidence.

Additional Remarks

We offer two additional comments on the Critique. First, the Critique is framed as a general call for rigor in the field, but is aimed bluntly at only one theory—STAGES. Again, perhaps as a Letter to the Editor, this is acceptable, but treating it as an article, its claim to these lofty ethical goals for "the field" comes into question because it does not mention any other theories, authors, or texts that exhibit the problems it claims to be addressing.^[19]

Second, the introduction to the Critique, "co-authored by leaders in the field," might give the impression that there is a substantial community or group behind the Critique. The number of co-signatories is in fact small, and it felt odd to us that Susanne, Beena, and Ken did not put themselves at the top of their critique as authors, but rather chose to distance themselves and frame the Critique as coming from a mysterious group ("we"), revealed only at its conclusion.

Conclusions and Questions from Nested Contexts

The “triple loop” action inquiry process developed by Bill Torbert and colleagues (2004, with Cook-Greuter as a co-author) invites us to inquire into and enact from several levels: the level of content (what is being investigated), the level of process (how we are thinking and inquiring), and the level of being (the consciousness and presence of those doing the inquiry). We have addressed each of the specific content issues in the Critique in turn, and have addressed more general issues noted in the Critique. We have stepped *in* a bit to report on our personal processes, likely limitations, and how we are “showing up” in this inquiry. We have stepped *back* in context to comment on the Critique article as a whole—is style and framing. We can step back one more level to look at the processes that has led to the publication of the Critique.

Given the amount of communication and interaction about the STAGES model by well-meaning people who have a history of collaboration, we have to ask: “how the hell did we get to this situation?” One in which our valued and esteemed colleagues (and friends) thought it necessary to draft a sweeping public condemnation. In our more reactive moments it is easy to notice imperfections in others that seem to have contributed. It is also humbling to Terri and those associated with Pacific Integral who counseled her along the way to inquire about how they might have contributed to this uncomfortable situation. All we will say here is that the reflective process is ongoing. Though we thought that a strong and direct Critique deserved a strong and direct Response, we continue to feel love and admiration for our colleagues and look towards healing any tensions that have emerged.

We want to again acknowledge that STAGES research is in its infancy, with just a tiny amount of empirical validation compared with Loevinger’s WUSCT. As to Esbjörn-Hargens’ reasonable call to specify “which aspects have been fully or partially validated and which ones are still speculative,” this is important yet we hope not to be held to a standard much higher than others in our community. We can say that what is validated is exactly what is described in the Research Overview, and any other statements about STAGES are in the realm of hypothesis or reasonable-but-unproven generalizations about the human condition—a category that includes much of integral theory writ large. Also, an overwhelming percentage of the claims, recommendations, and proclamations explicitly or implicitly made in our community about moving developmental research (or integral theory) *into practice* are not well supported by empirical research. We just don’t have a lot of evidence for how our assessments square with real-world behaviors, or what interventions help people or organizations become more developed, successful, or whole.[20] And such research is extremely expensive and time consuming, particularly for “independent scholars” and small businesses (most research reported in high level academic journals in the human sciences is funded by academic research projects costing \$100,000 to \$500,000).[21] In one sense we are all in the same boat here, and should certainly work together to increase clarity about what is and is not empirically tested.

Finally, we have stepped out further to reflect on the complexities of scholarly methodology, discussing the peer review process, tensions brought in by commercialization and idea-promotion. We hope that this minor flash-point in the history of the STAGES model, and in the history of models of meaning-making development, will contribute to the overall goals of knowledge building, relationship growth, and world-improvement. We appreciate this opportunity to respond and we have great respect for the authors of the Critique.

Future Directions: We conclude with a summary of some of what is planned for the STAGES model. A high priority of course is getting the validity research submitted in a peer reviewed journal. Other directions include:

- Unlike prior assessments in the Loevinger tradition, STAGES assessment is based on a domain-independent model of language structure (as opposed to being exemplar-based—explained more in the Companion Article). The underlying theory supports easy extension to **applications beyond the standard sentence completion** O’Fallon and others have begun to experiment with scoring other types of text, for example news articles, speeches, books, and social networking posts, for developmental levels. In such work it is the text (written performance) that is being scored, not the whole individual. This work is exciting but very preliminary, and as yet without validity studies.
- For similar reasons (domain-independence), the STAGES sentence completion test can be easily modified to include different stems, without the need to create new scoring manual chapters (which is the case for prior ego development assessments, and is quite time-consuming). Just as Torbert made moderate stem changes to include stems related to leadership contexts, O’Fallon has created “**specialty protocols**” including a set of stems targeted toward specific domains (so far these all have 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.
- An AI (machine learning) **computer-based scoring system** has been created for assessing developmental text based on the STAGES model (see StageLens.com). The automated scoring system is not yet as accurate as human scorers, but is accurate enough for group/aggregate studies and is already being used for several projects. The R&D on this project is ongoing.
- O’Fallon and her statisticians are **re-evaluating the OGIVE cutoff method** that has traditionally been used to aggregate the scores of the 36 sentence stems to produce the center of gravity score. They are investigating whether more modern statistical methods will yield more reliable total scores (or sub-scores).
- Two **dissertations** have been completed that make use of the STAGES model, and more are in process or being proposed. The eventual publication of these studies will increase the body of work using the STAGES model.

References

- Beck, D. E. & Cowan, C. C. (1996). *Spiral Dynamics*. Malden, MA: Blackwell Publishers, Ltd.
- Cohen, J. (1968). Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70, 213-220.
- 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.
- Commons, M. L. & Richards, F. A. (1984). A general model of stage theory. In M. L. Commons, F. A. Richards & C. Armon (Eds.), *Beyond formal operations: Late adolescent and adult cognitive development*, (pp. 120-141). New York: Praeger.

- Commons, M. L., & Pekker, A. (2008). Presenting the formal theory of hierarchical complexity. *World Futures: Journal of General Evolution* 64(5-7), 375-382.
- Cook-Greuter, S. (2013). Assumptions versus assertions: Separating hypotheses from truth in the integral community. *Journal of Integral Theory and Practice*, 8(3/4), 227.
- Cook-Greuter, S.R. (2013). *Ego Development: Nine levels of increasing embrace*. White paper available at www.cook-greuter.com.
- Kopko, K. C., Edwards, A., Krause, E., & McGonigle, V. J. (2016). Assessing Outcomes of National Science Foundation Grants in the Social Sciences. *Council on Undergraduate Research Quarterly*, 36(3).
- Darrall-Rew, J. & DiPerna, D. (2016). *Earth is Eden: An Integral Exploration of the Trans-Himalayan Teachings*. Integral Publishing House.
- Elster, J. (1999). *Alchemies of the mind: Rationality and the emotions*. Cambridge, UK: Cambridge University Press.
- Esbjörn-Hargens, S. (2009). An Overview of Integral Theory. An All-Inclusive Framework for 21st Century. Integral institute. Resource paper.
- Fischer, K. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological Review*, 87(6), 477-531.
- Fleiss, J.L. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York: John Wiley.
- Gura, T. (2002). Scientific publishing: Peer review, unmasked. *Nature*, 416(6878), 258-260.
- Haig, B. D. (1995). Grounded theory as scientific method. *Philosophy of education*, 28(1), 1-11.
- Hussein, M. E., Hirst, S., Salyers, V., & Osuji, J. (2014). Using grounded theory as a method of inquiry: Advantages and disadvantages. *The Qualitative Report*, 19(27), 1-15.
- Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Med*, 2(8), e124.
- Jefferson, T., Rudin, M., Brodney Folse, S., & Davidoff, F. (2006). Editorial peer review for improving the quality of reports of biomedical studies. The Cochrane Library.
- Kegan, R. (1994). *In over our heads: The mental demands of modern life*. Cambridge, MA: Harvard University Press.
- Lakatos, I. (1976). *Proofs and refutations: The logic of mathematical discovery*. J. Worrall & E. Zahar, (Eds.). Cambridge, MA: Cambridge Univ. Press.
- Lakoff, G. & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. New York, NY: Basic Books/Perseus Books Group.
- Landis, J.R.; Koch, G.G. (1977). "The measurement of observer agreement for categorical data". *Biometrics*. 33 (1): 159–174.
- Latour, B., 2005. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford: Oxford UP.
- 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.
- Mascolo, M. F., & Fischer, K. W. (2015). Dynamic development of thinking, feeling, and acting. *Handbook of child psychology and developmental science*. New York: John Wiley.
- Murray, T. (this volume, "the Companion Article"). On sentence completion assessments for ego development, meaning-making, and wisdom maturity, including STAGES.
- O'Fallon, T. (2013). The Senses: demystifying awakening. Paper presented at the Integral theory Conference, 2013, San Francisco CA.
- 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. (2011). stAGES: Growing up is Waking up—Interpenetrating Quadrants, States and Structures. Paper presented at the Integral theory Conference, 2011, San Francisco CA.
- Panksepp, J. (2005). Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and cognition*, 14(1), 30-80.
- Siegel, D. J. (2012). *Pocket Guide to Interpersonal Neurobiology: An Integrative Handbook of the Mind* (Norton Series on Interpersonal Neurobiology). WW Norton & Company.
- Torbert, B. & Associates (2004). *Action inquiry: The secret of timely and transforming leadership*. San Francisco: Berrett-Koehler.
- Torbert, W. (2011). The Practice of Action Inquiry. In P. Reason & H. Bradbury (Ed.s) *Handbook of Action Research*. Sage: London. 250-260.
- Wigglesworth, C. (2006). Why spiritual intelligence is essential to mature leadership. *Integral Leadership Review*, 6(3), 1-17.
- Wilber, K. (1995), *Sex, Ecology, Spirituality: The Spirit of Evolution*. Boston and London: Shambhala.

Wilber, K. (2005). *A sociable god: Toward a new understanding of religion*. Boston, MA: Shambhala Press.

Wilber, K. (2006). *Integral spirituality*. Boston, MA: Shambhala Press.

Appendix 1. A Memo from O'Fallon's Statistician

Comments on "Integral Theory Making and the Need for Empirical Rigor...."

Nayak Polissar, PhD

May 10, 2017

The Mountain-Whisper-Light Statistics

1827 23rd Ave. East, Seattle, WA 98112-2913

I and my colleague, Moni Neradilek, MS (Biostatistics), have been collaborating with Terri O'Fallon on an article on the STAGES method. The article concerns scoring using the STAGES method compared to scores that have been assigned by an independent scorer using the Loevinger/Cook-Greuter scoring method (hereafter, L/C-G). My statistical consulting firm (The Mountain-Whisper-Light Statistics) was responsible for the random selection of inventories (from a large pool of inventories) for this evaluation. We assigned the inventories to scorers in a manner that created a balanced representation of all the scorers; there were four scorers who used the L/C-G method and four scorers who used the STAGES method. All 16 combinations of L/C-G and STAGES scorers occurred in the resulting data, offering quite a diversity of scorers being compared between the two methods. We specified blinded scoring (a practice which the scorers followed), and we carried out the statistical analysis of the results. In fairness to Dr. O'Fallon, I do not wish to jeopardize the publication possibilities of the paper by quoting numeric results in advance, but I can offer some comments that should be helpful to this discussion.

First, the agreement is very strong between the assigned stage scores based on the STAGES method and those based on the L/C-G method. I have worked on many studies comparing methods for scoring or classifying people or phenomena, and the agreement in this study (based on a statistically substantial number of inventories) is very impressive.

We chose appropriate methods for evaluating agreement between the two scoring methods. The methodology used to support validation of STAGES in the article we are preparing is much more powerful method than the "interrater-reliability" measure cited in the Critique document. The O'Fallon study compares stages as scored by pairs of raters—each member of a pair using a different one of the two methods to rate the same inventory, blinded to each other's ratings. This exercise was carried out for many inventories, and the agreement between the two different methods was strong. An inter-rater comparison, by way of contrast, would have the two scorers using the same method for each inventory. A comparison of methods is far more informative (and more challenging) than a comparison of raters, and the STAGES method has done well in this more challenging exercise.

I have heard that there is some discussion that a correlation study is needed to validate a new method in comparison to an existing method. Pearson correlation has been brought into the discussion, as I understand it. However, correlation can be misleading in evaluating agreement between two methods. The classic 1983 paper by Bland and Altman (title: "Measurement in Medicine: the Analysis of Method Comparison Studies") has as its first conclusion: "Most common approaches, notably correlation, do not measure agreement." (Reference: D. G. ALTMAN and J. M. BLAND, *The Statistician* 32 (1983) 307-317.)

In the O'Fallon study we have used Cohen's Kappa to measure agreement. (Reference: Fleiss et al, *Statistical Methods for Rates and Proportions*, 2003.) Kappa is a far more meaningful (and demanding) measure of agreement than correlation. In our analysis, Kappa compares agreement (or disagreement) for each specific stage—a closer and more demanding comparison than correlation. A very simple illustrative example shows the problem that can arise in using correlation to measure agreement. Consider two scorers, independently scoring the same three protocols using two different methods, with results as follows.

Inventory completed by subject:	Stage, scored by method 1	Stage, scored by method 2
Joe	2.0	3.5
Mary	2.5	4.0
Ellen	3.0	4.5

The agreement between the two methods is, as you see, very poor. For example, the first subject who is being evaluated, Joe, is scored at stage 2.0 by the scorer using method #1 and at stage 3.5 by the scorer using method #2. The table shows that all three subjects have large disparities between the scores from the two methods. However, the Pearson correlation between the two methods is 1.0, the strongest possible correlation. A correlation of 1.0 indicates "perfect" correlation. It is obvious that perfect correlation does not mean perfect agreement. The value of Cohen's Kappa for these data is zero, appropriately. (Kappa = 1.0 means perfect agreement, Kappa = 0.0 means no agreement.)

If, on the other hand, the inventories of Joe, Mary and Ellen were each scored with an exactly matching stage by the two methods, the correlation would still be 1.0 but Kappa would now be 1.0, appropriately. The problem with the correlation for this purpose is evident. Correlation does not measure agreement. That is why this study used the more demanding agreement statistic, Cohen's Kappa, which showed strong agreement between the scores based on the STAGES method and the scores based on the L/C-G method. Cohen's kappa is very widely used in agreement studies.

Finally, I understand that there is some concern, based on considerations from developmental theory, about the STAGES method. I am not conversant with developmental theory, so I do not know how to evaluate whether or not Dr. O'Fallon's theoretical justification

for her method fits well with existing theory or with methods for developing new theory. I do know that she has something that works. The study results will show that STAGES scoring does well in matching up with the Loevinger/Cook-Greuter scoring method. That agreement alone should carry considerable weight in evaluating her method. If I may pose a homey analogy that bears on this discussion, consider a track coach who has a novel method for training her runners. Her runners consistently improve their times and move up in the rankings. However, someone suggests that the method is faulty because it is too different from current training methods. Well, there is the famous phrase, “the proof is in the pudding.” The runners using the method are running faster. The method works. Perhaps the new method represents an expansion of the field.

As an additional note, I understand that Dr. O’Fallon does not consider this article as the last step in the development of STAGES. While it will be a very worthy contribution to the literature, the article will suggest some next steps in testing and further development of this methodology. The article will also have some other features that I have not mentioned here, such as an evaluation of some proposed newly designated stages. It certainly deserves a chance at a hearing. And, hopefully, other researchers will compare the Loevinger/Cook-Greuter scoring method with STAGES and in their own studies.

I hope that this is a helpful contribution to the discussion.

Footnotes

[1] <http://www.verticaldevelopment.com/maturity-assessment-for-professionals.html>

[2] One reason was that Terri spent over a year tending full time to a dying family member. Also, Terri enlisted the services of two senior statisticians for her research design, analysis, and publication co-authorship. Collaborating with these (busy) professionals has been extremely valuable, but also introduced significant delays.

[3] Terri inferred this from several things: Ken recently wrote a testimonial and agreed to write a forward for Terri’s in-process book about STAGES; he agreed to let her use his graphics; and he cites O’Fallon’s work, along side of Cook-Greuter’s, in his recent book *The Religion of Tomorrow*.

[4] In one sense STAGES turns some AQAL patterns on their heads, while not being in conflict with AQAL. AQAL categorizes all reality in one map, and developmental lines happen within each quadrant of that map (and radiate out into all quadrants simultaneously). STAGES flips this ordering, and puts the quadrant/zone map *within* a developmental sequence (repeated with each of its three tiers).

[5] We must note several things here. First, of course, all developmental theories describe general patterns and therefore reduce the complexity of the human condition to a small number of key concepts. Thus, such theories can never be expected to make concrete predictions about any individual or situation, but rather are meant to hold more generally. Ego (or meaning making) development is also a “thick” developmental line, encompassing many sub-capacities that may be developing at different rates, and estimates of a person’s ego development, while taking these into account, are simplifications in this sense as well.

[6] Also, as is usually the case with alternative models, the definitions of terms, or precisely what they point to, overlap but are not identical. (For example, when Jung talked about the unconscious his meaning of the term was probably a bit different than Freud’s. To acknowledge this is to take a construct-aware perspective on knowledge building and scientific inquiry [compatible with theories by Lakoff & Johnson, 1999; Lakatos 1976; and Elser 1999].) Though Wilber has made a major contribution in helping the field differentiate stages and stages and inquire into their relationship, he does not own the definition of the terms, which are general terms used broadly. Also, O’Fallon recognizes the problems with finding the right terms and does grapple with how to name and define ideas adequately.

[7] Darrall-Rew & Diperna report that “it is likely the case that radical awakening [state-stage/vantage point stabilization] becomes a prerequisite for most advances [structure stages of development and mastery on Earth and in the kosmos... In his more advanced work on integral spirituality, Ken Wilber has suggested that the structures of evolutionary development (1st, 2nd, and 3rd tier) each partake in a particular state or vantage point. For him the 1st and 2nd tier structures...all involve awareness situated in the Gross state but as 3rd tier structures unfold, one gains access to evolutionary altitudes that begin to objectify the gross state of the kosmos and then the Subtle state and then the Causal state. ...the actualization of Supermind...requires the recognition of Nondual Awakened Awareness, but the recognition of nondual Awakened Awareness does not require the actualization of Supermind.” (pgs. 119 & 120).

[8] STAGES divides the MAP Diplomat level into two second-person perspective levels, divides that MAP Construct Aware level in two fifth-person perspective levels, and divides the MAP Unitive level into two sixth-person perspective levels—all based on the theoretical model.

[9] “Orienting generalization” is also a term Wilber uses for his overall style of scholarly argumentation, which tries to synthesize what is generally agreed upon by contemporary scholarship (though some critique this method for over-generalization, ignoring important counter-theories, and other limitations meta-narrative). But sticking to the AQAL quadrant “orienting generalizations,” Esbjörn-Hargens (2009, p. 4; and see Wilber, 2006): says that “...there are at least two ways to depict and use the quadrant model... the [quadratic approach] highlights four irreducible dimensions that all individuals have and quadrivia refer to the four fundamental perspectives that can be taken on any phenomena. In either case, the four quadrants or quadrivia are co-nascent... and are mutually implicated in one another.”

[10] One possible area of confusion is that, though STAGES is not a theory of everything, it is a theory about how people make meaning about everything (i.e. anything), and draws from the same categorical distinctions. This relates to Wilber’s distinction about quadrants and “quadrivia,” the former being ontological categories and the latter being perspectives (epistemological descriptors). STAGES coordinates the quadrivia with the ego development framework.

[11] If we consider the dimensions STAGES uses to measure (score) text to infer developmental level as orienting generalizations,

then one could say that STAGES does measure orienting generalizations in a sense. But for now we assume that by orienting generalizations the authors mean the quadrants/zones in use as ontological categories rather than perspective (epistemological) markers.

[12] Note: the use of state-related vantage points is a newer aspect of the scoring method, that was not fully integrated in the method used 4 years ago when the primary research study data was gathered.

[13] O'Fallon employed the most qualified statisticians she could find (using criterion co-developed with Elliott Ingersol, one of the co-signatories of the Critique). Polossar has a PhD in statistics from Princeton and has published over 200 peer reviewed scientific articles, and his co-analyst has an MA in statistics and has published over 80 peer reviewed scientific articles. Both are prodigious meditators (one for over 40 years), and have incredibly high ethical standards.

[14] We should also note that only one of the four trained scorers was also previously trained in Cook-Greuter's scoring method. The statistical analysis was done with and without this scorer included to insure that this factor was not significant.

[15] For examples one can search Google or Google Scholar for "periodic table psychological" or "periodic table social." In addition, O'Fallon, upon hearing Cook-Greuter's complaint on this matter about a year ago, has not been using this metaphor.

[16] 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."

[17] 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."

[18] See http://publicationethics.org/files/Ethical_guidelines_for_peer_reviewers_0.pdf .
<http://icmje.org/recommendations/browse/roles-and-responsibilities/responsibilities-in-the-submission-and-peer-review-process.html>,
and <http://icmje.org/recommendations/browse/roles-and-responsibilities/responsibilities-in-the-submission-and-peer-review-process.html>.

[19] As one example, we can point to the research on Spiritual Intelligence done by Cindy Wigglesworth (2006). This important work faces the same challenges as the STAGES research—it is groundbreaking, there are very few studies so far, there is a profit making business attached to it, the promising conclusions of studies are used to promote the business in ways that might seem too far-reaching from the perspective of "objective" academic scholarship, and it re-uses many of the concepts of integral theory and adult development in novel ways. Yet, rather than include this study within the discussion of rigor in the field, Susanne has aligned herself with this work and praises it. (We agree that it is excellent work.)

[20] For example, Loevinger and Cook-Greuter describe dozens of attributes of the personality or capacity at each of their developmental levels. For example see Table 1 in the Companion Article—this set of attributes for each level are not empirically validated in any detail. We don't have strong evidence for most of these attributes individually, and they are given almost by definition for what is expected to be found at each level.

[21] Kopko et al., 2016.

About the Authors



Terri O'Fallon is a researcher, teacher, coach, spiritual director and designer of transformative containers. She does ongoing research on the Integral STAGES developmental model which supports a MetAware tier with two later levels of development.

Terri is a partner of Developmental Life Design, which creates programs based on the STAGES model. She also is a partner in Pacific Integral, which creates transformational programs in later level Leadership. Terri holds Masters degrees in Special Education, in Spiritual direction and an Integral PhD in Transformative Learning and Change.



Tom Murray, Ed.D., is a Senior Research Fellow at the University of Massachusetts School of Computer Science and is Chief Visionary and Instigator at Perspexity Solutions. His projects include an R&D project for virtual scoring of ego-development testing; and research on supporting “social deliberative skills” and deep reflective dialogue in online contexts. He is an Associate Editor for Integral Review journal, and he has published many articles on integral theory as it relates to education, contemplative dialog, leadership, ethics, knowledge building communities, epistemology, and post-metaphysics. Email: tommurray.us@gmail.com. Web: www.perspexity.com, www.tommurray.us.



Geoff Fitch is a coach, trainer, and facilitator of growth in individuals and organizations, and a creator of transformative leadership education programs worldwide. He is a founder of Pacific Integral, where he was instrumental in the development of the internationally-acclaimed Generating Transformative Change program, now offered three continents and in its 24th cohort. Through these programs, he has researched and developed novel approaches to individual and collective growth, and has designed and facilitated dozens of residential learning retreats. He has been exploring diverse approaches to cultivating higher human potentials for over 25 years, including somatic and transpersonal psychology, mystical traditions, innovation and creativity, leadership, integral theory, and collective intelligence. Geoff also has over 30 years experience in leadership in business. He holds a master's degree in Transpersonal Psychology from Sofia University and B.S. in Computer Science, magna cum laude, from Boston University, and has additional studies jazz music, philosophy, and management. Learn more at www.pacificintegral.com and www.geoff-fitch.com.



John Kesler lives in Salt Lake City, Utah and divides his time in four general ways. He practices and shares integral polarity practice, a practice developed by John, grounded in mindfulness which works with polarities throughout the spectrum of one's being and the breadth of one's life, relationships and organizations. He is a social activist who leads a non-profit called the Salt Lake/Global Civil Network which does integrally informed social and political transformation work locally and networks globally. John is a practicing commercial transactional attorney, and he spends a lot of time with his wife, Colleen (Freddie), his two children and their families. John does consulting, teaching and writing primarily in the first two areas mentioned in ways that are deeply interpenetrating.



Kim Barta M.A. is an internationally recognized licensed professional psychotherapist, coach, spiritual guide, and speaker. His work and insights spring from grounded experiential practice with self and others in his cross cultural and life long experiences.

Kim worked for gender rights in the small rural school in central Montana. In college, Kim founded Students in Union with Nature, which partnered College students with alternative energy advocates in the community creating a solar powered organic garden lifestyle for college students.

Kim graduated with a B.A. in Cultural Anthropology from the University of Montana in 1984. He pursued a Masters in an interdisciplinary program that combined Psychology, Social Work, Counseling, and Sociology in a broad perspective for healing. He

lived with a Native American Shaman and did his Masters Thesis in Shamanism and Modern Psychotherapy. He graduated with high marks in 1986.

After developing and implementing a successful construction training program for people suffering from severe mental illness: Bi-polar, Schizophrenia, and Axis II Disorders, Kim pursued outdoor leadership, becoming an instructor with the International Outward Bound schools. Kim went on to develop a mental health treatment program for children and youth that was adopted by 3 hospitals in Montana. In 1992, Kim founded and continues to co-operate a mind/body healing arts clinic in Polson, Montana, nestled beneath the high mission mountains on the shore of Flathead Lake, in the Flathead Indian nation.

Ever seeking better ways of healing in the life of self and others, families and children, Kim, has developed several successful new forms of therapy that deeply and rapidly lead to healing. These have proven themselves effective with a wide range of issues including: Depression, Anxiety, PTSD, Anger, Addiction, Chronic Pain, physical illnesses and weight loss and personal growth. People utilize his psychology/coaching/spiritual guidance practice, globally.

Currently, Kim has teamed up with Dr Terri O'Fallon to present workshops and trainings internationally in a new model of human development designed and researched by Dr. O'Fallon. Kim's developmental learning theory fits lock and key with Dr O'Fallon's Stages model and his extensive history in clinical human dynamics provides a grounded passion to the workshop experience.

- [Tweet](#)
- [Facebook](#)
- [LinkedIn](#)