Measuring Altitude

Empirically Grounding Developmental Levels with the Lectical Scale

Brendan Graham Dempsey

Director of Research, Institute of Applied Metatheory

June 2026

INSTITUTE OF APPLIED METATHEORY

appliedmetatheory.org

Contents

Executive Summary

540+	5	.85	7
PERFORMANCES SCORED	DEVELOPMENTAL MODELS	PEAK CORRELATION (R)	ALTITUDE BANDS ANCHORED

For a quarter century, Integral Theory has attempted to organize the field of adult developmental psychology around a bold claim: that the many stage models charted by researchers such as Lawrence Kohlberg, Cheryl Armon, Robert Kegan, Jane Loevinger, and James Fowler are not rival accounts of human growth but parallel expressions of one underlying developmental gradient. Ken Wilber called this gradient *altitude* and rendered it as the familiar color spectrum running from Infrared through Turquoise and beyond. Yet the correlations behind the altitude charts were assembled *conceptually*—by aligning stage descriptions that merely sounded alike—rather than *empirically*, by measuring the same texts with a common metric. Critics inside and outside the integral community have long noted that this leaves the construct vulnerable: an elegant meta-conjecture awaiting operationalization and testing.

This white paper presents the results of a multi-year research program that puts the altitude construct to exactly that testing. Drawing on more than 540 interview performances spanning five canonical developmental models, we scored original research data—including Kohlberg's own Moral Judgment Interviews, Cheryl Armon's Good Life interviews, a substantial portion of James Fowler's original *Stages of Faith* sample, and others—using the Lectical Assessment System, Theo Dawson's calibrated, domain-general measure of hierarchical complexity in the neo-Piagetian tradition of Fischer and Commons.

The findings are unambiguous. Stage assignments in every model correlate strongly and significantly with independently measured hierarchical complexity: $r = .85$ for moral development ($n = 403$, $p < .001$), $r = .82$ for evaluative reasoning ($n = 61$, $p < .001$), $\rho = .72$ for faith development ($n = 54$, $p < .001$), and $\rho = .60$ for ego development ($n = 20$, $p = .006$), with the small Kegan sample ($n = 5$) tracking the same trend. Moreover, the regression lines for the different models converge on a shared developmental corridor: equivalent stages across models cluster at equivalent Lectical scores, anchoring the altitude map's color bands, for the first time, to quantitative positions on a validated scale.

The result is an updated, empirically grounded altitude map in which modal Magenta centers near Lectical 875, Red near 950, Amber near 1025, Orange near 1100, Green near 1175, Teal near 1200, and Turquoise near 1250. This constitutes a successful first-pass validation of one of Integral Theory's central meta-conjectures—that the developmental "lines"

share a common structural "altitude" dimension—while also disciplining the construct: altitude, on this account, is hierarchical complexity expressed in domain-specific meaning-making, measurable to within a quarter of a developmental level. The paper closes with implications for integral research practice, limitations of the present datasets, and the next phases of IAM's measurement agenda.

Key findings at a glance

Moral development (Kohlberg & Walker), $n = 403$	$r = .85$
Evaluative reasoning (Armon), $n = 61$	$r = .82$
Faith development (Fowler), $n = 54$	$\rho = .72$
Ego development (Loevinger/Cook-Greuter), $n = 20$	$\rho = .60$
Orders of consciousness (Kegan), $n = 5$	*trend*

Every estimable coefficient falls in the strong range ($p < .001$; ego development $p = .006$). Equivalent stages across models cluster at equivalent Lectical scores, anchoring the altitude map's colour bands to positions on a validated scale.

1 Introduction: From Meta-Theory to Measurement

Integral Theory is a metatheory. Where ordinary scientific theories take real-world phenomena as their data, a metatheory takes *theories themselves* as its data, providing a higher-order map of an entire knowledge domain. This is both its great power and its standing methodological liability. As Mark Edwards has argued, metatheory building must be complemented by metatheory testing: the conjectures generated at the metatheoretical level must be operationalized into hypotheses, methods, and measurements that can be verified against middle-range theory and, ultimately, against empirical phenomena (Edwards, 2013). Without this descending arc of the research cycle, metatheoretical claims remain interesting speculations—synoptic pictures that organize the literature but cannot be confirmed, refined, or falsified by it.

Few metatheoretical conjectures have been more consequential for the developmental field—or more conspicuously untested—than Ken Wilber's construct of *developmental altitude*. In *Integral Psychology* (2000), Wilber synthesized over one hundred developmental models into a single comparative chart, proposing that stage sequences as different as

Kohlberg's moral stages, Loevinger's ego stages, and Fowler's stages of faith all ascend a common structural gradient. He later rendered this gradient as a color spectrum—the altitude markers Infrared, Magenta, Red, Amber, Orange, Green, Teal, Turquoise, and beyond—so that "similar" stages across models could be aligned to the same color band (Wilber, 2006). The altitude construct underwrites the AQAL framework's treatment of levels and lines: distinct developmental lines (cognition, morality, self-identity, values, faith) are understood to traverse the same vertical scale, like separate paths up one mountain.

The construct's vulnerability has always been how the alignments were made. The correlations in the integral charts were established by comparing the *content* of stage descriptions across the published literature—a procedure whose consequences Sara Ross (2008) memorably likened to playing "telephone games" with developmental theory, in which conceptual resemblance substitutes for structural measurement and errors compound as alignments are passed from chart to chart. Two stage descriptions can sound alike while organizing experience at quite different levels of structural complexity, and vice versa. What the altitude conjecture required, in other words, was a common metric: a single, calibrated, domain-general yardstick against which performances from any developmental line could be scored, so that cross-model alignment becomes an empirical finding rather than an editorial judgment.

Two stage descriptions can sound alike while organizing experience at quite different levels of structural complexity.

Zachary Stein and Katie Heikkinen (2008) identified precisely such a yardstick. They argued that the Lectical Assessment System (LAS) operationalizes key aspects of Wilber's altitude construct, offering "a common metric" for syncing up lines of psychological development. The proposal was programmatic: the LAS *could* validate altitude, if anyone gathered the data. This white paper reports what happens when someone does. Drawing on the empirical research program documented in Dempsey (2026), *Psyche and Symbolic Learning*, the second volume in the *Evolution of Meaning* book series, we present Lectical analyses of original interview data underlying five canonical developmental models—Kohlberg and Walker's moral development research, Armon's evaluative reasoning studies, Kegan's constructive-developmental theory, Loevinger/Cook-Greuter's ego development framework, and Fowler's faith development theory—comprising over 540 scored performances in total. The question is simple: *When the same complexity ruler is laid against all of these models at once, do their stages in fact line up the way the altitude map predicts?*

The answer, we will show, is *yes*—to a degree that is frankly striking for first-pass validation work in the human sciences. The remainder of this paper proceeds as follows. Section 2

reviews the altitude construct and its validation problem. Section 3 introduces hierarchical complexity and the Lectical Scale as the common metric. Section 4 describes the datasets and method. Section 5 presents results model by model, and Section 6 synthesizes them into an updated, empirically anchored altitude map. Sections 7 and 8 discuss limitations and chart the road ahead for IAM's research agenda.

2 The Altitude Construct and Its Validation Problem

Within the AQAL model, "altitude" names the degree of developmental complexity available to consciousness at a given level, abstracted from any particular line. Just as inches can measure the height of mountains, buildings, or trees without being "made of" any of them, altitude is meant as a content-free vertical metric applicable across the lines: a person can be at one altitude in cognition, another in moral reasoning, another in emotional intelligence, yielding the familiar integral "psychograph." The color scheme gives the metric its taxonomy: Magenta names magic-impulsive organization, Red egocentric-opportunistic organization, Amber mythic-conformist, Orange rational-achievement, Green pluralistic, Teal and Turquoise integral, and so on (Wilber, 2006).

Key terms

Altitude. Wilber's content-free vertical measure of developmental complexity, shared across all lines and rendered as a colour spectrum from Magenta to Turquoise.

Hierarchical complexity. The formal order at which actions coordinate and transform the actions below them; the structural backbone the developmental models are shown to share.

Lectical Scale. Theo Dawson's calibrated instrument scoring the hierarchical complexity of a performance, from Lectical 700 to 1399.

AQAL. Wilber's integral framework, organising development into *levels* (altitude) and *lines* (domains such as morality, self, and faith).

Stated this way, altitude is a hypothesis about latent structure. It predicts that if performances from different developmental lines were measured on a single structural scale, (1) each model's stage sequence would correlate strongly with the scale, and (2) the stage-to-scale mappings would *converge* across models—Kohlberg's Stage 3, Fowler's Stage 3, Kegan's 3rd order, and Loevinger's Conformist stage should occupy approximately the same region of the scale, because all are expressions of the same underlying level of complexity in different domains of meaning-making.

Notably, this is a risky prediction in the Popperian sense, and the stakes for Integral Theory are high. Cataloguing the recurring themes in how people reason about justice or the good might yield a serviceable *typology*—a set of qualitatively distinct meaning-making styles—without yielding a genuinely *developmental* model, since nothing in a typology guarantees that its categories are structurally ranked rather than merely different. Establishing that one has an ordered sequence of increasing complexity, rather than a flat catalogue of worldview flavors, requires probing the underlying structural dimension directly, with a measure independent of the content being classified. Moreover, even if each model did exhibit an overall developmental trend, that would not guarantee that the *specific* stages align across models, such that one model's third stage sits at the same structural level as another's. Until recently, the appropriate measurement infrastructure simply did not exist.

Notably, nearly 20 years ago, Ross (2008) emphasized that the Commons-Richards Model of Hierarchical Complexity, filed in integral charts as merely the "cognitive line," is in fact "a content-free, domain-independent general theory" applicable to every line—that is, a candidate for the altitude metric itself rather than one line among others. Stein and Heikkinen (2008) drew the practical conclusion: the Lectical Assessment System, built on the hierarchical complexity construct, "uses one method to assess development in many lines," and could therefore operationalize altitude for integral researchers. What follows takes up that program.

3 The Common Metric: Hierarchical Complexity and the Lectical Scale

The neo-Piagetian tradition resolved a long-standing problem in stage theory by separating the *structure* of development from its *content*. Michael Commons's Model of Hierarchical Complexity (MHC) defines a sequence of orders of complexity in purely formal, task-analytic terms: actions at each order coordinate and transform the actions of the order below, in a non-arbitrary way, such that each new order constitutes a genuinely higher-level organization rather than a mere chain of lower-level elements (Commons, 2008). Because the definition is content-free, the orders apply to any reasoning domain—logical, moral, interpersonal, spiritual. Kurt Fischer's dynamic skill theory arrived at a parallel sequence of levels through the empirical study of skill construction, including the characteristic clustering of developmental spurts and the dependence of performance on context and support (Fischer, 1980).

Theo Dawson and colleagues turned this convergent theoretical structure into a calibrated psychometric instrument. The Lectical Assessment System (LAS) scores the hierarchical complexity of reasoning performances—typically interview transcripts or written

responses—on a continuous scale in which each major level occupies 100 points. The portion of the scale relevant to post-infancy human development runs from Level 7 (representational mappings, roughly early childhood) through Level 13 (principled mappings, the rarefied upper reaches of adult reasoning), corresponding to Lectical scores from 700 to 1399. Decades of validation research demonstrate high inter-rater reliability, strong internal consistency, and—crucially for present purposes—domain-generality: the same scoring system applies to performances about morality, education, leadership, energy, etc. (Dawson, 2002; Dawson-Tunik et al., 2005; Stein & Heikkinen, 2008).

Table 1. The Lectical Scale: levels, names, and score ranges relevant to this study (after Fischer and Dawson).

Lectical Level	Fischer/Dawson Name	Score Range	cf. Piaget	cf. Commons (MHC)
8	Representational Systems	800–899	Concrete (early)	Concrete
9	Single Abstractions	900–999	Concrete–Formal (early)	Abstract
10	Abstract Mappings	1000–1099	Formal	Formal
11	Abstract Systems	1100–1199	Formal (late)	Systematic
12	Single Principles	1200–1299	Post-formal	Metasystematic
13	Principled Mappings	1300–1399	Post-formal	Paradigmatic

Two features of the scale matter for what follows. First, it is fine-grained: a quarter-level phase (25 points) is a meaningful and reliably detectable unit, which means cross-model alignments can be tested with real resolution rather than at the coarse grain of whole stages. Second, it is independent of all the stage models under test. Lectical scores track the structural organization of a text without reference to Kohlberg's, Fowler's, or anyone else's stage definitions. When a Moral Judgment Interview and a Faith Development Interview both receive a score of 1100, that agreement is produced by a measurement procedure blind to the hypothesis being tested—which is exactly what validation requires.

4 Data and Method

The analyses reported here score original interview performances from the primary research traditions behind each developmental model. Wherever possible, we obtained the actual historical data—Kohlberg's own Moral Judgment Interviews, Fowler's original *Stages of Faith* sample, etc.—so that the validation bears on the models as their authors built them,

not on later reconstructions. All texts were submitted to Lectica, Inc. for hierarchical complexity scoring using the Lectical Assessment System; stage scores derive from the original researchers' published assignments. Table 2 summarizes the corpus.

Table 2. Datasets analyzed, by developmental model.

Model	Source Data	n	Notes
Moral development (Kohlberg)	Kohlberg's Moral Judgment Interviews	105	Original MJI transcripts from Kohlberg's research program
Moral development (Walker)	Walker longitudinal MJI corpus	298	Family-based longitudinal study (Walker, 1989)
Evaluative reasoning (Armon)	Good Life interviews	61	17-year longitudinal/cross-sectional study of "the good"
Self development (Kegan)	Exemplar SOI texts, CDT scoring manual	5	Collated exemplar passages plus one full interview; illustrative only
Ego development (Loevinger/Cook-Greuter)	WUSCT scores paired with FDI transcripts	20	Interview serves as proxy performance for Lectical scoring
Faith development (Fowler)	Original Stages of Faith sample (1972–1981)	54	38 full interviews, 10 excerpts, 6 composites; >20% of Fowler's n = 359
Faith development (ongoing)	Bielefeld–Chattanooga FDI studies (2002–2024)	172+	Rescoring in progress; preliminary results reported

Statistical analysis paired each performance's stage assignment with its Lectical score. Pearson correlations were computed for models whose stage variables can be treated as interval-like (moral and evaluative reasoning, where half-stage scoring yields a quasi-continuous variable); Spearman rank correlations were used where ordinal treatment is more defensible (faith stages; the small ego-development sample). Following the conventions of the source research program, scatterplots display continuous stage scores, while boxplots round to the half-stage to make level clustering visible. All figures in Section 5 are reproduced from the source analyses in Dempsey (2026).

5 Results

5.1 Moral Development (Kohlberg)

Lawrence Kohlberg's six-stage model of moral judgment provides the largest dataset in this study. Combining Kohlberg's own interviews (n = 105) with Lawrence Walker's longitudinal corpus (n = 298) yields 403 scored performances spanning moral Stages 1 through 5. The correlation between moral stage and Lectical score is very strong and highly significant: $r = .78$ ($p < .001$) for the Kohlberg data, $r = .87$ ($p < .001$) for the Walker data, and $r = .85$ for the combined sample (Figure 1).

The two datasets, collected decades apart by different researchers, trace practically identical trends—the Walker regression line is only slightly steeper. Restricting the analysis to performances with the highest Lectical confidence ratings only strengthens the relationship further, to $r = .80$ for the Kohlberg subset (n = 55) and $r = .89$ for the Walker subset (n = 190).

Rounding the continuous stage scores to half-stages reveals the level clustering predicted by the hierarchical complexity account (Figure 2). Stage 1 performances begin at Lectical Level 8 and ascend into early Level 9; Stage 2 occupies upper Level 9 into lower Level 10; Stage 3 sits predominantly in mid-to-late Level 10; Stage 4 moves into Level 11; and the available Stage 5 data cluster in late Level 11. Linear complexity alone predicts approximately 72% of the variance in moral development level. The movement from preconventional through conventional to postconventional moral reasoning is, demonstrably, a complexification process.

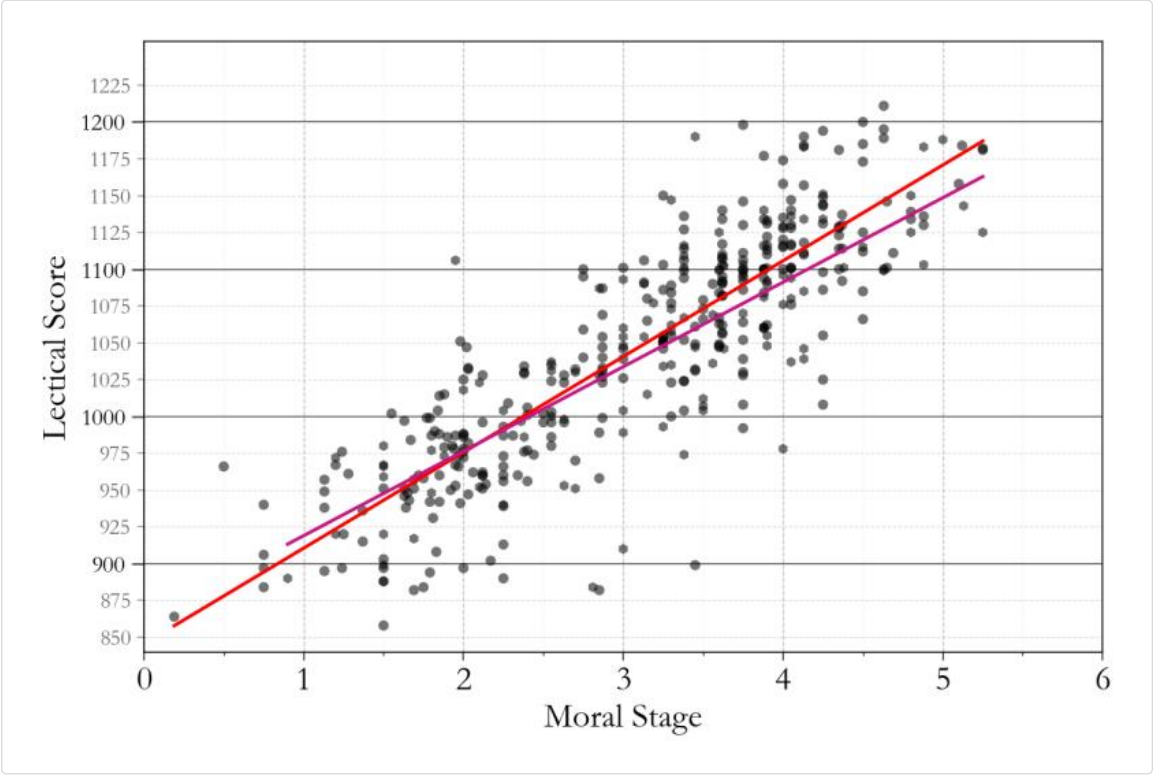


Figure 1. Moral stages by Lectical score (n = 403). Red line: Walker data; magenta line: Kohlberg data.

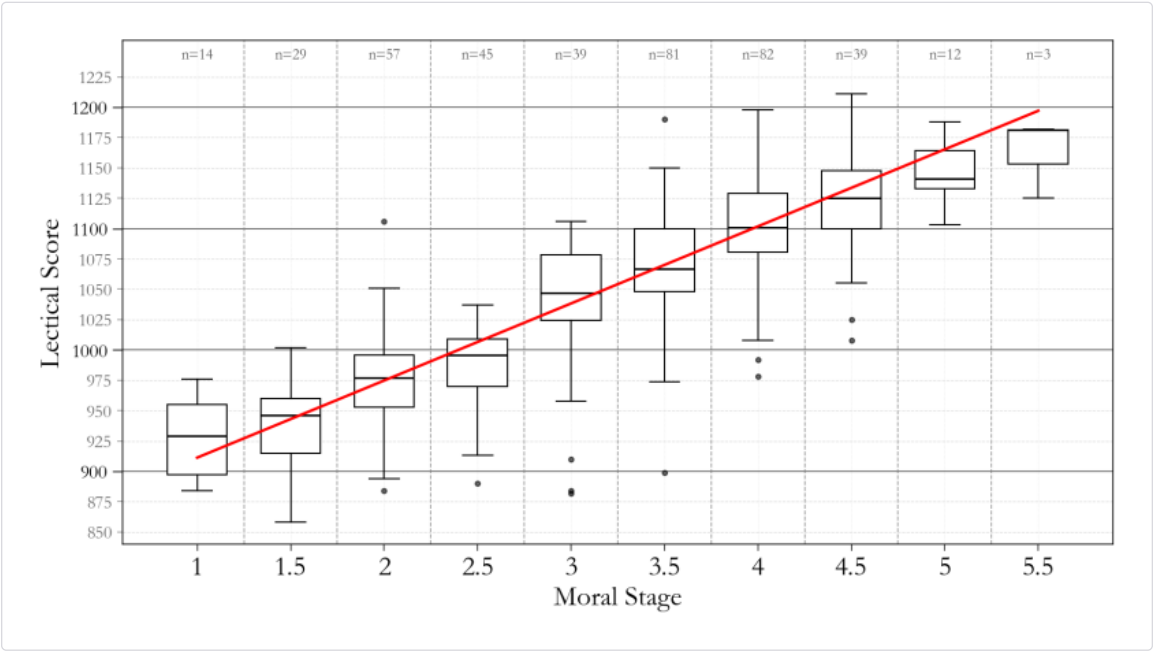


Figure 2. Moral Stages 1–5 by Lectical score, rounded to half-stages (n = 401).

5.2 Evaluative Reasoning (Armon)

Cheryl Armon, a student of Kohlberg's, conducted the first systematic longitudinal study of how conceptions of "the good"—the good life, good work, good friendship—develop across the lifespan. Armon identified five stages across the developmental spectrum through which such evaluative reasoning evolves (with gestures to a tentative sixth). She refers to these stages as 1) Egoistic Hedonism, 2) Instrumental Hedonism, 3) Altruistic Mutuality, 4) Individuality, and 5) Autonomy/Community. The five stages of Armon's bear a direct similarity to Kohlberg's stages—indeed, they exhibit clear overlaps in terms of thematic content. Are both tapping the same deep construct of hierarchical complexity?

Using a sample of 61 Good Life interviews scored by Lectica, the correlation between evaluative reasoning stage and Lectical score is again very strong: $r = .82$, $p < .001$ (Figure 3).

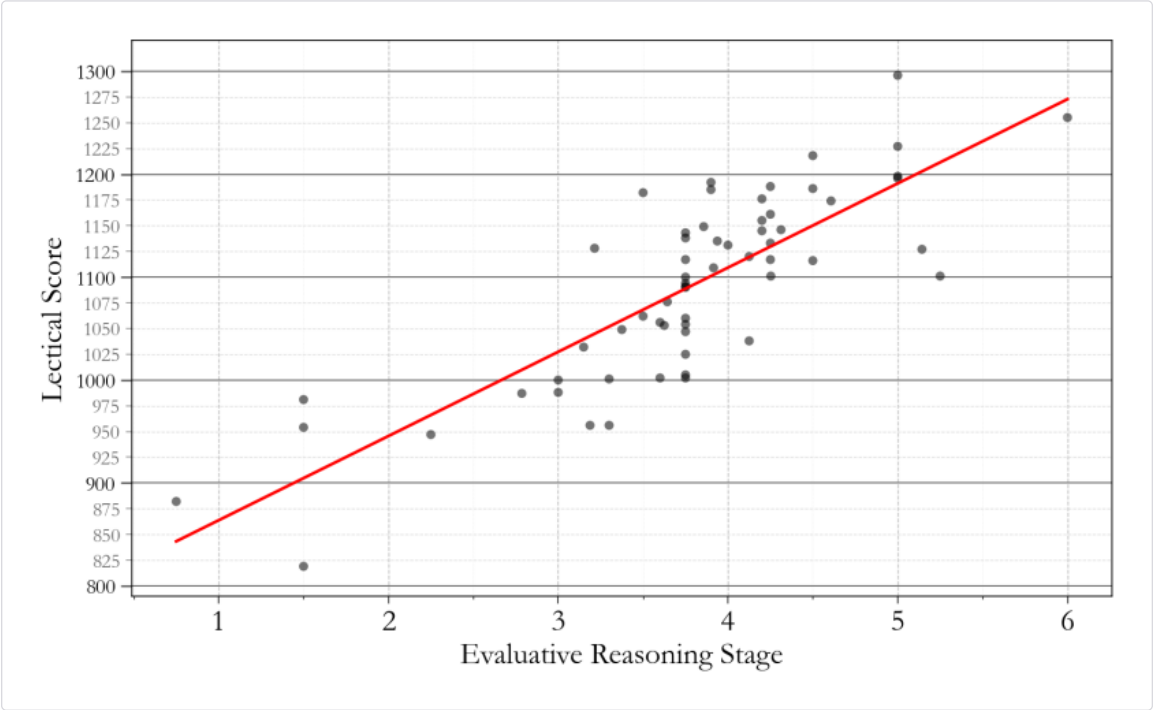


Figure 3. Value stages by Lectical score (n = 61).

Armon's stage-to-level mapping reproduces the Kohlberg pattern (at slightly steeper slope): Stage 1 at late Level 8/early Level 9, Stage 2 in upper Level 9, Stage 3 predominantly in Level 10, Stage 4 predominantly in Level 11, and Stages 5–6 breaking decisively into Level 12 (Figure 4).

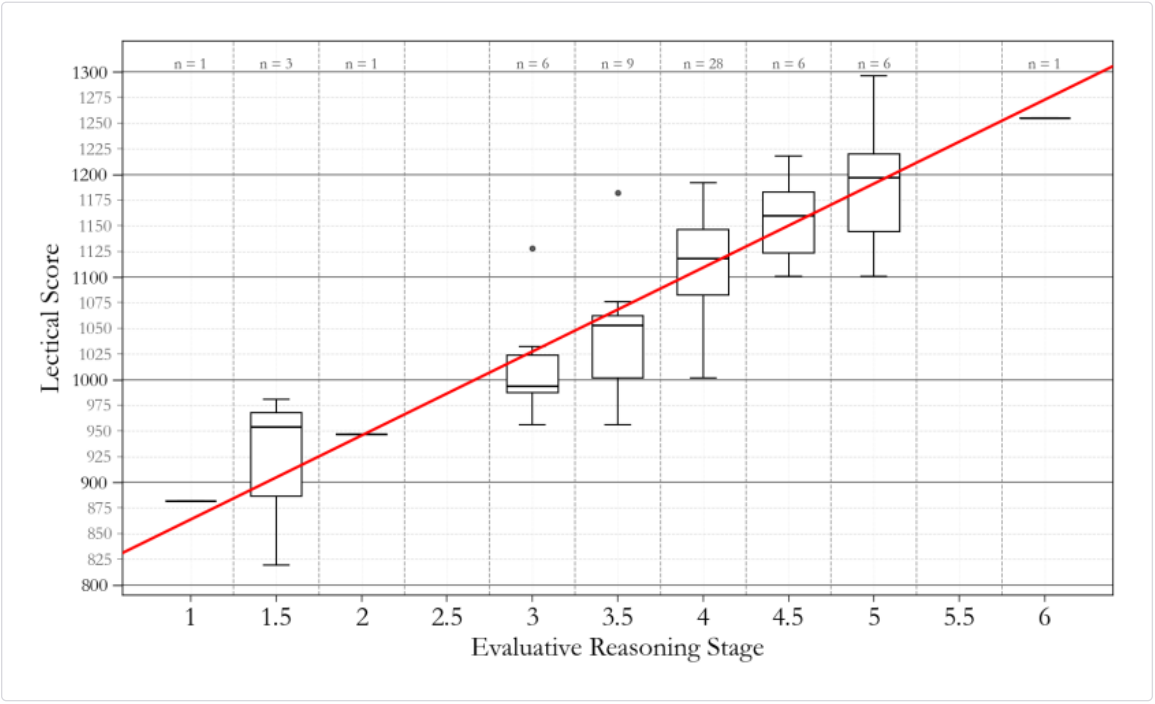


Figure 4. Value Stages 1–6 by Lectical score, rounded to half-stages (n = 61).

The same structural transformations visible in moral judgment thus appear in valuation more generally. Armon's model can be understood, then, as a domain-specific stage theory of evaluative reasoning measurable in terms of the Lectical Scale. And, as Kohlberg's stage model theoretically unfolds across the same ranges of complexity, the models would be charting the same developmental territory.

5.3 Orders of Consciousness (Kegan)

Robert Kegan's Constructive-Developmental Theory (CDT) casts ego development as successive transformations in the subject-object relationship: what one *is subject to* at one order becomes an *object* one can reflect upon at the next. He traced this process across five "orders of consciousness." Original Subject-Object Interview data proved difficult to obtain, so the analysis here relies on the exemplar texts in the official SOI scoring manual: 24 short passages collated by order of consciousness into scorable texts, plus one full interview. With n = 5 scored texts the sample is far too small for statistical inference, but the results are worth reporting for their consistency with the larger datasets (Figure 5): the collated 1st-order texts scored at Level 8, 2nd-order at Level 9, 3rd-order at Level 10 (the full 3rd-order interview at 1150), and the 4th-order texts at 1050. The contours of the same complexification gradient are visible.

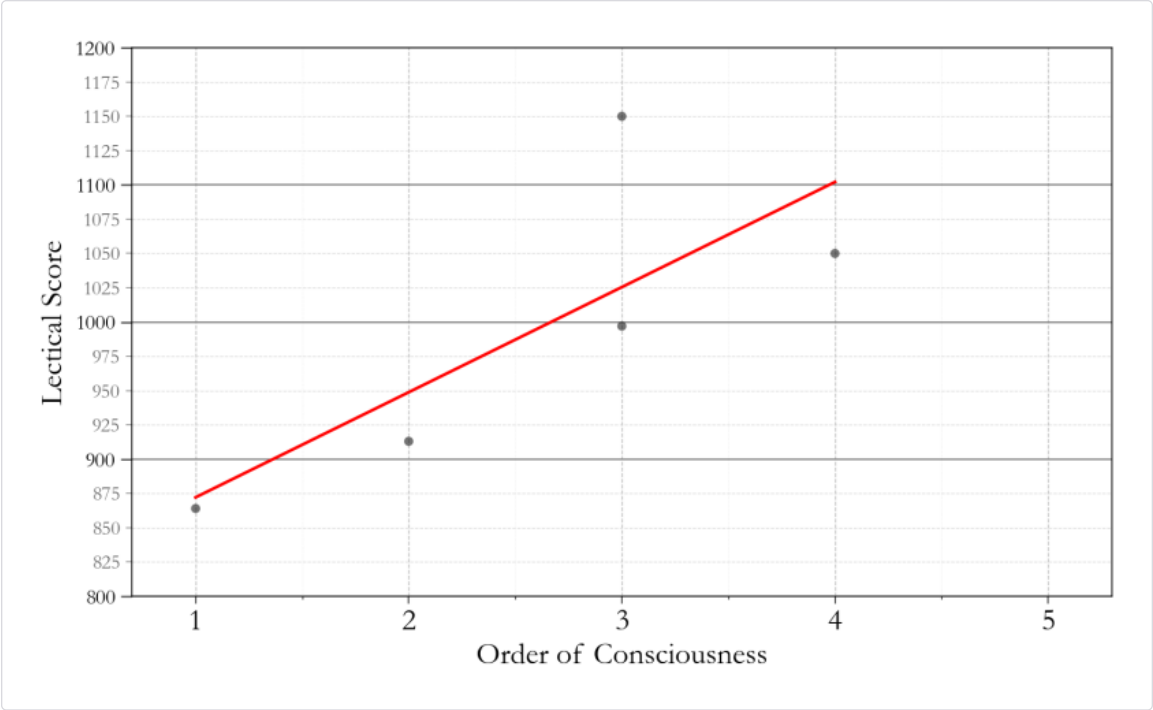


Figure 5. Orders of consciousness by Lectical score (n = 5). Illustrative only.

5.4 Ego Development (Loevinger and Cook-Greuter)

Jane Loevinger's Ego Development Theory (EDT), extended by Susanne Cook-Greuter, emerged from psychometrics rather than the Piagetian clinical tradition, and is measured by the Washington University Sentence Completion Test (WUSCT) rather than by interview. This methodological gulf has made structural comparison with the other models notoriously difficult; correlations offered in the literature (including Kegan's own EDT–CDT alignments) have rested on surface similarities between stage descriptions. The present study exploits a rare opportunity: within the faith development dataset, 20 interviewees also completed the WUSCT, providing an EDT ego stage and a Lectically scorable interview from the same person. Treating the interview as a proxy performance, ego stage correlates significantly with Lectical score: $\rho = .60$, $p = .006$ (Figure 6). The data are noisy and the sample small, but the complexification trend is clear, and the stage placements are consistent with the hypothesized crosswalk to Kegan's orders.

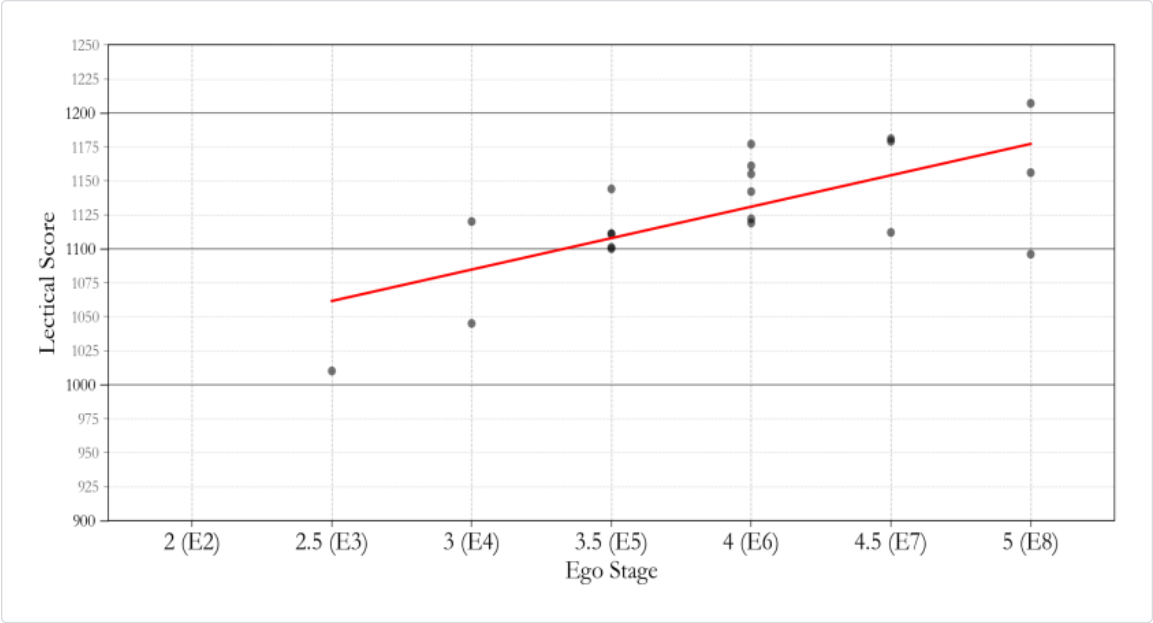


Figure 6. Ego development (EDT) stage by Lectical score of associated FDI transcript (n = 20).

5.5 Faith Development (Fowler)

James Fowler's Faith Development Theory addresses the broadest construct of all: faith—not as religious belief per se, but as the human activity of meaning-making with respect to ultimate concern. For integral audiences, Fowler's stages anchor the "spiritual intelligence" line of the altitude charts. We obtained a substantial portion of the original sample on which *Stages of Faith* (1981) was built—38 full interviews, 10 interview excerpts, and 6 composite interview passages, together representing over 20% of Fowler's 359 cases—and submitted them for Lectical scoring.

Faith stage correlates strongly with hierarchical complexity: $\rho = .72$, $p < .001$ (Figure 7), closely in line with prediction (Dempsey & Kurze, 2025). Stage 1 clusters in Level 8, Stage 2 in Level 9, Stage 3 largely in Level 10 (reaching into Level 11), Stage 4 in Level 11, and the highest stages press toward Level 12.

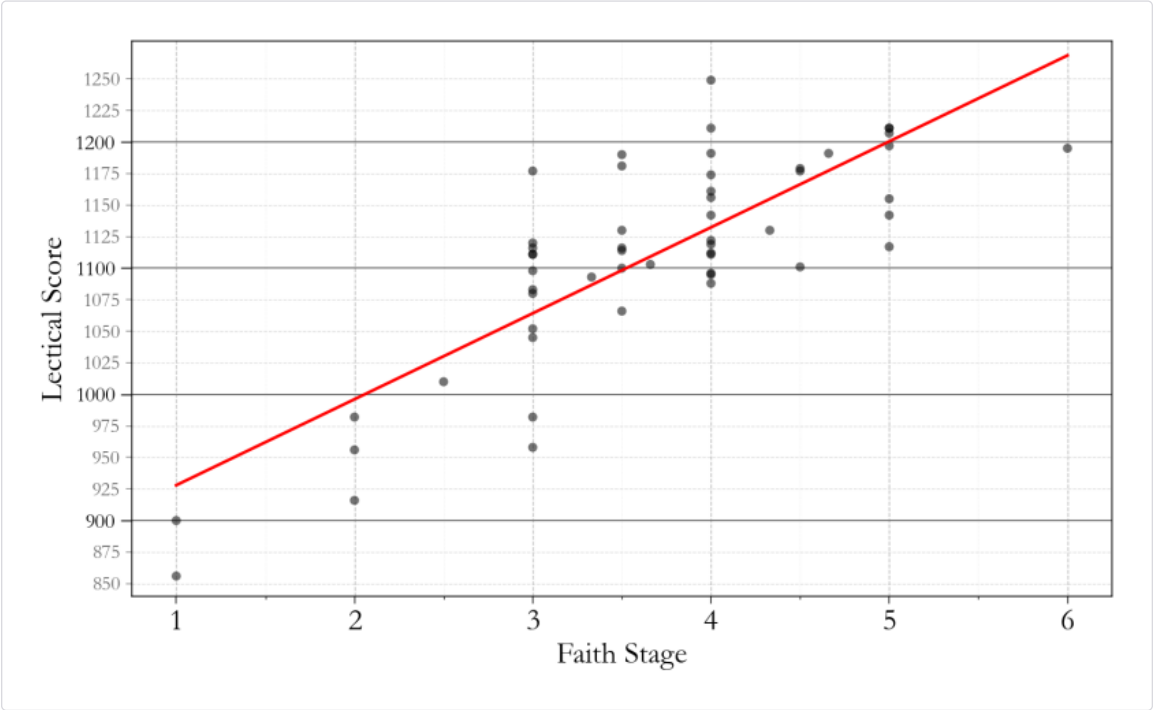


Figure 7. Faith stages by Lectical score, original Fowler sample (n = 54; interviews 1972–1981).

Work is ongoing with a far larger corpus of over 500 Faith Development Interviews from the Bielefeld–Chattanooga research program (2002–2024). These have been scored for hierarchical complexity, and systematic recoding for faith stage is underway. Preliminary results from my 172 rescored interviews show an even tighter relationship (Figure 8)—as expected, since structure can now be measured objectively on both axes rather than inferred from content. This dataset, the largest faith-development corpus ever assembled for structural analysis, will anchor the next phase of validation.

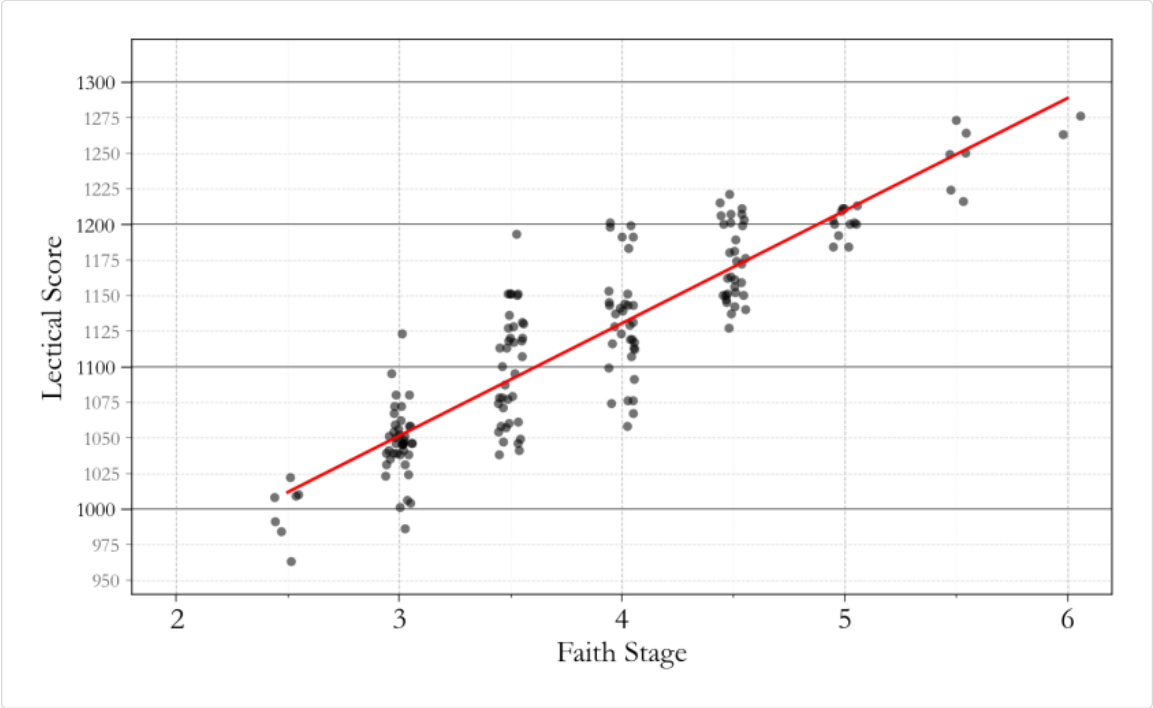
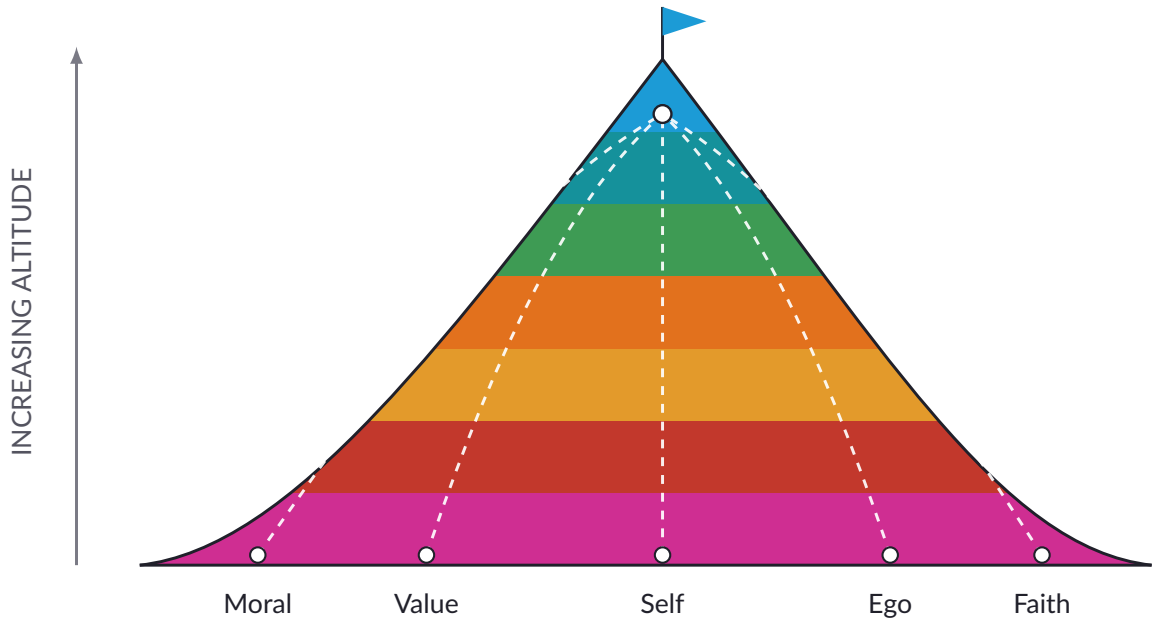


Figure 8. Faith stages by Lectical score, Bielefeld–Chattanooga sample, preliminary (n = 172; interviews 2002–2024).

5.6 Combined Analysis: One Mountain, Many Paths



One mountain, many paths. Each developmental line begins in a different domain of meaning, yet all ascend the same complexity gradient; the summit marks the highest measured altitude. A schematic of the paper's central finding, borne out quantitatively in Figure 9.

The decisive test of the altitude construct is not any single correlation but the pattern across all of them. Figure 9 plots all 543 performances from the five models on common axes, with each model's regression line overlaid.

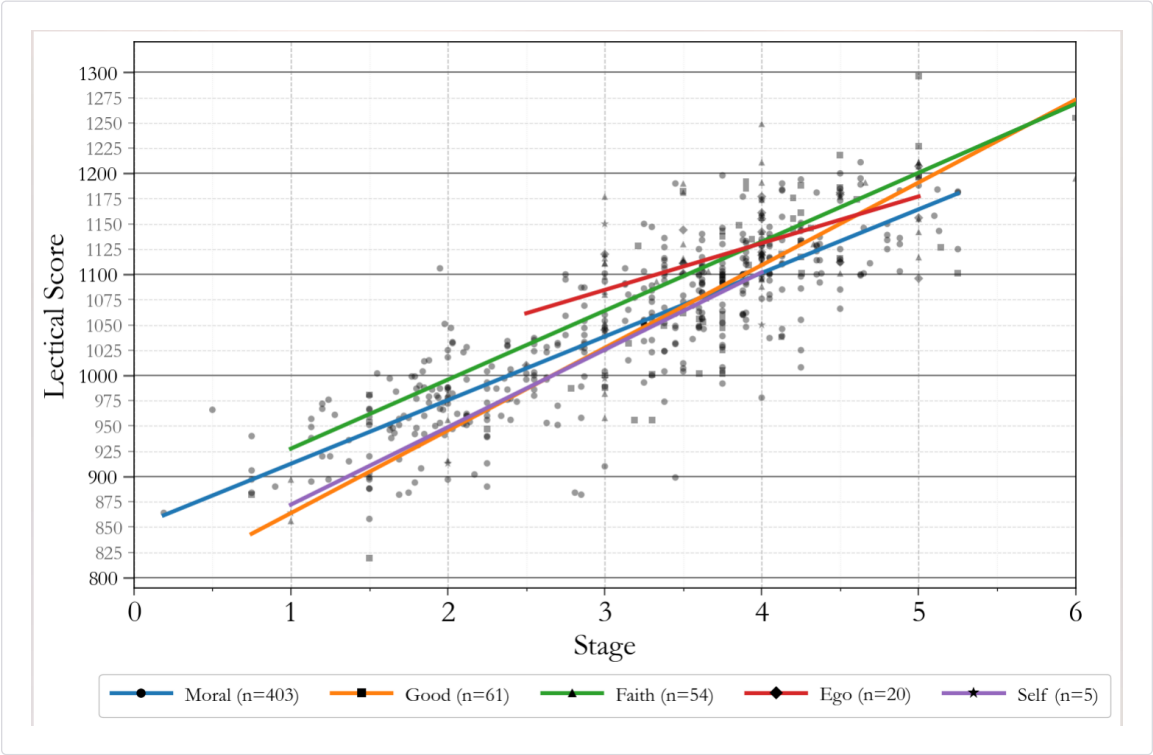


Figure 9. All data: stage by Lectical score across five developmental models (total n = 543), with per-model regression lines; presentation of combined analyses.

Two features stand out. First, every line slopes steeply upward: in every domain, higher stages mean higher hierarchical complexity. Second—and this is the altitude prediction proper—the lines run through a shared corridor. A Stage 3 performance lands near Lectical 1025–1075 whether the interview was about a moral dilemma, the good life, or God; a Stage 4 performance lands near 1100–1150 across the board. The models do not merely each correlate with complexity; they converge on a common stage-to-complexity mapping.

A Stage 3 performance lands near Lectical 1025–1075 whether the interview was about a moral dilemma, the good life, or God.

Figure 10 summarizes the correlation evidence. By the conventions of the field, every estimable coefficient falls in the strong range or at its threshold, with the largest samples

producing the strongest relationships—the signature of a real effect emerging from measurement noise as power increases, rather than an artifact inflated by small samples.

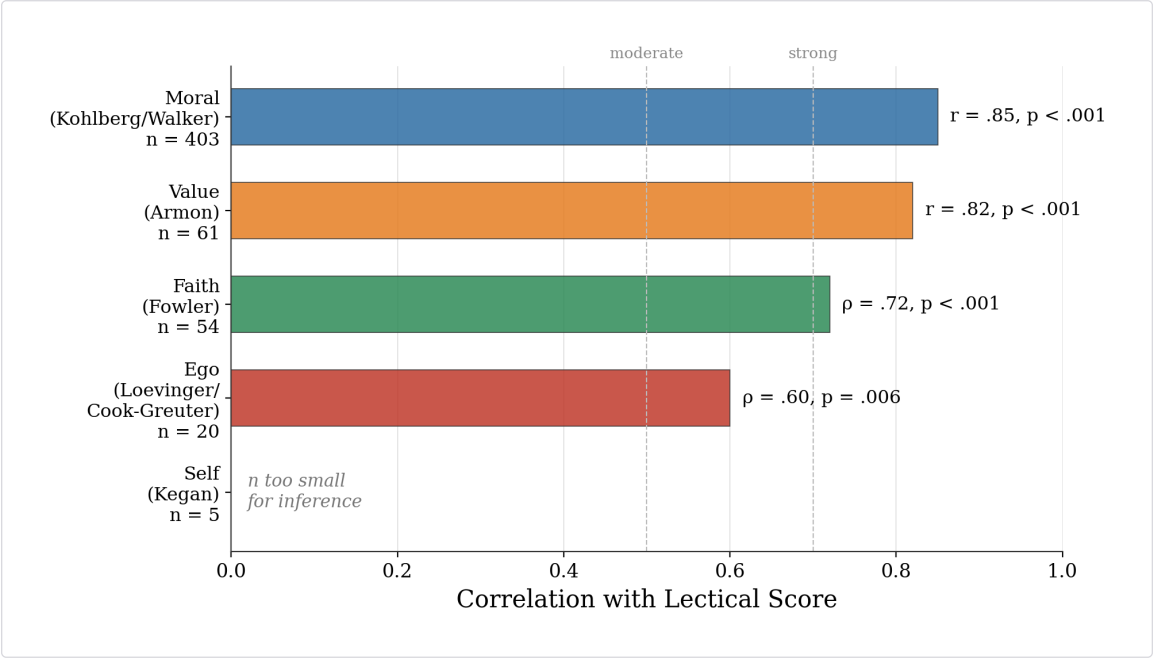


Figure 10. Correlation between stage assignment and Lectical score, by developmental model. The Kegan sample (n = 5) is consistent with the trend but too small for inference.

6 An Empirically Grounded Altitude Map

These convergent results allow us to do something the integral community has wanted for twenty-five years: anchor the altitude colors to quantitative positions on a validated scale. Because equivalent stages across models cluster at equivalent Lectical scores, the qualitative color bands can be assigned score anchors, and every model's stages can be placed on the spectrum by *measurement* rather than by interpretive alignment of stage descriptions.

Figure 11 presents the resulting updated chart, restricted—deliberately—to those models for which Lectical data exist. Entries in parentheses denote stages that are theoretically expected at a given altitude but for which scored data are not yet available (e.g., Kohlberg's Stage 6, Kegan's 5th order, Cook-Greuter's Unitive stage).

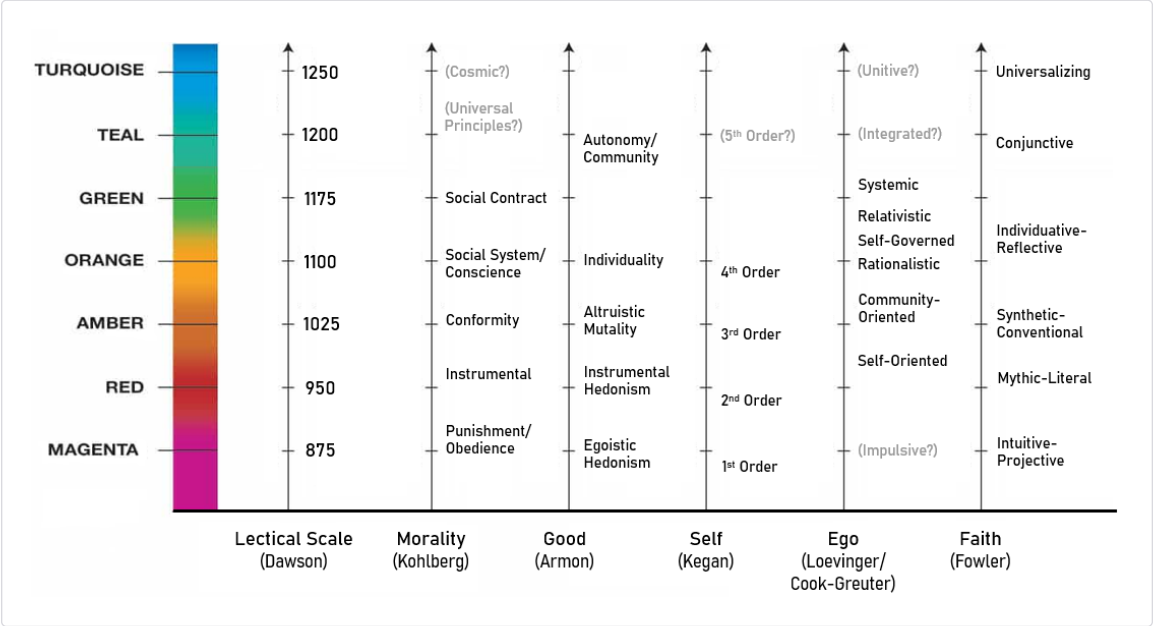


Figure 11. The updated, empirically grounded altitude map: the altitude colour bands anchored to Lectical score ranges, with each model's stages placed by measurement rather than interpretive alignment. Parenthetical entries are theoretically expected but not yet scored.

7 Discussion and Limitations

What exactly has been validated? The central finding is that the canonical models of meaning-making development—moral, evaluative, self, ego, and faith—share a latent structural dimension, and that this dimension is hierarchical complexity as operationalized by the Lectical Scale. This vindicates the core of the altitude conjecture: there really is one structural "mountain" that the various developmental "paths" ascend, and it can be surveyed with a common instrument. At the same time, the results discipline the construct. Altitude is not some occult property of consciousness; on this account it is the measured level of hierarchical organization a person brings to meaning-laden tasks. Complexity is not the *whole* story of any of these models—in the moral data it accounts for roughly 72% of variance, leaving real room for domain-specific content, affect, and context—but it is demonstrably the structural backbone they share.

Several limitations deserve emphasis. First, sample sizes vary enormously across models. The moral findings rest on 403 performances and are robust; the Kegan analysis rests on 5 collated exemplar texts and is purely illustrative; the ego-development analysis ($n = 20$) uses an interview as a proxy performance for a construct normally measured by sentence completion, which introduces method variance even as it cleverly bridges the two tradi-

tions. More SOI and paired WUSCT data are the clearest research need. Second, the upper reaches of the spectrum remain thinly evidenced everywhere: there were no Stage 6 moral interviews in Kohlberg's data or ours, no scored 5th-order Kegan texts, and only a single Universalizing faith case—so the Teal and Turquoise anchors carry wider error bars than the conventional-range anchors. Third, a conceptual caution: the correlation of "whole-self" models (CDT, EDT) with a skill-based complexity metric should not be read as reinstating global stage theory. Following the neo-Piagetian consensus, we interpret such results in terms of a person's typical functional range of complexity across overlapping self-relevant domains, not as a monolithic mental structure that progresses in lockstep. The altitude map describes statistical regularities in performance, not a ladder the whole psyche climbs one rung at a time.

8 Implications and the Road Ahead

For integral theory and practice, the immediate implication is methodological. Researchers and practitioners who invoke altitude—in coaching, organizational development, education, or spiritual direction—can now anchor that talk to a calibrated scale with published reliability, rather than to chart alignments of uncertain provenance. Cross-model translation ("where does a Fowler Stage 4 sit relative to a Loevinger E6?") becomes an empirical lookup rather than a judgment call. And the psychograph, integral theory's signature applied tool, gains what it has always lacked: a common vertical axis on which different lines can be plotted with genuine metric meaning, just as Stein and Heikkinen envisioned in 2008.

Cross-model translation becomes an empirical lookup rather than a judgment call.

For the developmental research community, the findings suggest a program of consolidation. Rather than treating Kohlberg, Armon, Kegan, Loevinger, and Fowler as competing paradigms, they are better understood as domain-specific lenses on a single complexification process—what Dempsey (2026) theorizes as *Symbolic Learning*, the recursive construction of ever more hierarchically complex meaning structures in human culture. Each model contributes irreplaceable domain insight (the phenomenology of faith is not the phenomenology of justice reasoning), while the Lectical Scale supplies the structural spine that relates them. This is metatheory functioning as Edwards prescribed: the meta-conjecture

descends through operationalization to measurement, and returns validated and refined.

Within IAM's portfolio, this white paper is the keystone of a broader measurement agenda. The Faith Development Pathway is building on the Fowler validation and the Bielefeld–Chattanooga rescoring to produce the world's first open-access, empirically calibrated protocol for the maturation of ultimate concern. The Cultural Complexity Index extends the same Lectical methodology from ontogeny to history, scoring 5,000 years of religious, philosophical, and scientific texts to test whether collective meaning-making exhibits the same complexification signature at societal scale—the question of psychosocial development to which the individual-level findings reported here naturally lead. And the Worldview Studies Initiative is generating large-scale data on the distribution of developmental worldviews in living populations. Together these projects aim at a genuinely cumulative science of the noosphere: individual and collective development, from metatheory to measurement.

A quarter century ago, *Integral Psychology* offered the developmental field a magnificent hypothesis drawn in colored bands. The data are now in a position to repay that ambition.

The bands are real; the mountain has a height; and we can measure it.

Claude Fable 5 was used to organize and distill these findings into the present form of this white paper.

References

Armon, C. (1984). Ideals of the good life: A longitudinal/cross-sectional study of evaluative reasoning in children and adults. Doctoral dissertation, Harvard University.

Armon, C., & Dawson, T. L. (1997). Developmental trajectories in moral reasoning across the life span. Journal of Moral Education, 26(4), 433–453.

Binder, T. (2023). Ego Development for Effective Coaching and Consulting. Heidelberg: Carl-Auer.

Colby, A., & Kohlberg, L. (1987). The Measurement of Moral Judgment (2 vols.). Cambridge: Cambridge University Press.

Commons, M. L. (2008). Introduction to the Model of Hierarchical Complexity and its relationship to postformal action. World Futures, 64(5–7), 305–320.

Cook-Greuter, S. R. (1999). Postautonomous ego development: A study of its nature and measurement. Doctoral dissertation, Harvard University.

Dawson, T. L. (2002). New tools, new insights: Kohlberg's moral judgement stages revisited. International Journal of Behavioral Development, 26(2), 154–166.

Dawson-Tunik, T. L., Commons, M., Wilson, M., & Fischer, K. W. (2005). The shape of development. European Journal of Developmental Psychology, 2(2), 163–195.

Dempsey, B. G. (2026). Psyche and Symbolic Learning (The Evolution of Meaning, Vol. 2, 2nd ed.). Stannard, VT: Sky Meadow Press.

Dempsey, B. G., & Kurze, H. (2025). Hierarchical complexity as a domain-general framework for faith development. Sky Meadow Institute Working Paper. http://dx.doi.org/10.2139/ssrn.5275680

Edwards, M. G. (2013). Misunderstanding metatheorizing. Systems Research and Behavioral Science, 31(6), 720–744.

Fischer, K. W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. Psychological Review, 87(6), 477–531.

Fowler, J. W. (1981). Stages of Faith: The Psychology of Human Development and the Quest for Meaning. San Francisco: Harper & Row.

Kegan, R. (1982). The Evolving Self: Problem and Process in Human Development. Cambridge, MA: Harvard University Press.

Kegan, R. (1994). In Over Our Heads: The Mental Demands of Modern Life. Cambridge, MA: Harvard University Press.

Kohlberg, L. (1981/1984). Essays on Moral Development (Vols. 1–2). San Francisco: Harper & Row.

Lahey, L., Souvaine, E., Kegan, R., Goodman, R., & Felix, S. (2011). A Guide to the Subject-Object Interview: Its Administration and Interpretation. Cambridge, MA: Minds at Work.

Loevinger, J. (1976). Ego Development: Conceptions and Theories. San Francisco: Jossey-Bass.

Ross, S. N. (2008). Using developmental theory: When not to play telephone games. Integral Review, 4(1), 31–46.

Stein, Z., & Heikkinen, K. (2008). On operationalizing aspects of altitude: An introduction to the Lectical™ Assessment System for integral researchers. Journal of Integral Theory and Practice, 3(1), 105–139.

Walker, L. J. (1989). A longitudinal study of moral reasoning. Child Development, 60(1), 157–166.

Wilber, K. (2000). Integral Psychology: Consciousness, Spirit, Psychology, Therapy. Boston: Shambhala.

Wilber, K. (2006). Integral Spirituality: A Startling New Role for Religion in the Modern and Postmodern World. Boston: Integral Books.

About

About the Author

Brendan Graham Dempsey is a Director of Research at the Institute of Applied Metatheory. He's also a writer, poet, farmer, and the director of Sky Meadow Institute, an organization dedicated to promoting systems-based thinking about the things that matter most. He holds a BA in religious studies from the University of Vermont and a master's in religion and art from Yale University. He is the author of the 7-volume Metamodern Spirituality Series and, most recently, Metamodernism: Or, The Cultural Logic of Cultural Logics. His primary interests include theorizing developments in culture after postmodernism, productively bridging the divide between science and spirituality, and developing sustainable systems for life to flourish. All of these lead through the paradigms of emergence and complexity, which inform all of his work.

About the Institute of Applied Metatheory

The Institute of Applied Metatheory (IAM) is an international applied philosophy network dedicated to the education and application of "big picture" philosophical systems known as integrative metatheories. In conjunction with the nonprofit IAM Foundation, IAM provides scholars, practitioners, and organizations with the resources and support they need to advance integrative metatheory and apply it to promising evolutionary leverage points for human flourishing in the 21st century. Learn more at appliedmetatheory.org.