

Deepening Our Understanding of Developmental Assessments Use in Developing Leaders' Capacity for Complexity

Aiden Thornton. (2023). *Facing the Complexity Gap: Developing Leaders' Reasoning Skills to Meet the Complex Task Demands of their Roles*. University of Western Australia.

Review by Jonathan Reams¹

In the 25 years or so that I've been engaged in the field of developmental psychology, and more specifically as it is applied to leadership, I have observed both my own maturation of understanding the field as well as how the discourse in the field itself is maturing. Part of this process is noticing and participating in how the general practitioner discourse evolves. While this is useful to see and generally healthy, (less mature, overly simplistic applications of these models has never been helpful), at times more empirical work is needed to help clarify distinctions and claims being made in the field. It is in this context that I was eager to read through Aiden M. A. Thornton's doctoral dissertation.

I have followed Aiden's research project for many years, having had early conversations with him 10 years or more ago, hearing about how he was setting up his tools for doing the work, being involved in a client project with him, and participating in the public seminars based on his findings three years ago. (You can see the slides from these public seminars [here](#)). All of this has given me a healthy respect for Aiden and the quality of his work.

In this review, I intend to do three things. One is to provide a relatively robust summary of what I see as the key points in Aiden's research. My hope is that this summary will broaden the range of people who can benefit from encountering this work. Not everyone likes to read hundreds of pages of dissertation material, but many practitioners in the field who work with adult development theories and measures do have an interest in evolving their own practice and being informed about current updates, distinctions and contextualization.

A second intention is to provide some specific commentary on aspects of Aiden's research. This will include observations and perspectives that have come up for me along the way of reading through the dissertation. There are many points worth discussing and clear positions Aiden is taking that are worth digging into. The third intention is to zoom out and briefly reflect on my perception of the implications this work has for the field as whole.

I would also like to say something about my own familiarity with the two main models that Aiden examines. My early exposure in the mid '90s to the field of adult development came through

¹ **Jonathan Reams** is Editor-in-Chief of Integral Review, holds a position at the Norwegian University of Science and Technology, is co-founder of the Center for Transformative Leadership and does consulting and leadership development work for a variety of clients in Europe and North America.

Jonathan@Reams.com



values models (e.g., Brian Hall's Values Shift), ego development (e.g., Bill Torbert and Susann Cook-Greuter) and Robert Kegan's work. It was not until about 2008 that I became aware of Theo Dawson and Zak Stein's work with Lectica and was exposed to Kurt Fischer's dynamic skill theory. I have made attempts to provide simple introductions to the key distinctions between these models, for example in [this blog post](#). I also wrote [this piece](#) on the history of developmental theory, where I tried to address what I perceived as a lack of awareness of Fischer's skill theory in the communities of discourse I encountered. Finally, I edited [this academic anthology](#) aimed at showcasing the richness and diversity of adult development work in the field of leadership.

Opening Comments

Aiden's dissertation takes aim at some hard questions. We all talk about the complexity of challenges being faced today, yet it is not as clear as we might wish that the models we use to address that complexity are actually fit for that purpose. Within the broadly defined field of adult development, there are many theories, models, frameworks, and assessments and many of them have been applied directly to leadership and its development. There is a significant amount of research from a few individuals that provide a foundation for the larger body of consultants and coaches who apply these in practice. Yet few practitioners are able to take the time and energy required to thoroughly address sets of assumptions inherent in their use of these theories, models, frameworks and assessments. Resonance with these leads many practitioners to make a more casual use of them. While we can say that any model or assessment in the hands of a good practitioner can be of value for a client, it is worth stepping back and questioning this. This is one way I would characterize the impulse behind Aiden's research.

Aiden invites us into exploring what he has found from asking difficult questions with deep rigor, to create a more "intellectually spacious conversation" (p. 303) for the field as a whole. As I reflect on this, I wonder if it is just that we have better tools now to examine these questions? Or is it also possibly that not so many of us have been interested enough in the challenges these questions pose to do the work necessary to adequately address such questions? This work is not for the faint hearted and requires the use of highly technical statistical and mathematic tools.

I also want to present the punchline up front before summarizing the lengthy journey involved in Aiden coming to his findings. In my own words, the take-away is that while complexity is a central theme of leadership today, there are significant differences between how Lectica's assessments address complexity and how sentence completion tests in ego development models approach complexity. Even though both theories and assessments use the term complexity, ego development theories "may not be well suited to sense-making about a particular type of complexity" (p. 25). The distinction about what type of complexity is meant here relates to a second key point.

Since the 1980s, when Kohlberg and Loevinger were debating the application of Piaget's criteria for what qualifies as a 'stage,' there has been a desire to meet what are considered 'hard stage criteria' in order to justify the claim that stages are sequenced in an ascending order of hierarchical complexity. Hard stage models are defined by five criteria: unidimensionality, invariant sequencing, qualitative distinctness, structured wholeness, and hierarchical integrations. Aiden's through examination of these criteria led to the conclusion that scores yielded by the

Lectical Assessment System (LAS) have provisionally been shown to meet hard stage criteria, while scores yielded by ego development scoring system only meet one of the five criteria, the one for being invariantly sequenced. This positions them as ‘soft stage models.’

Aiden is clear that being positioned as a soft stage model in itself does not mean that ego development models and measures are problematic. He states that they are important ways to look at and understand human experience. What he makes clear is that using the Sentence Completion Test to make inferences about leaders’ ability to navigate complexity can be problematic, and potentially using ego development theory to help leaders make sense of complexity may also be problematic even if ego development measures such as the SCT are not used. Sentence completion assessments do provide data. The question is what claims do we make about that data and how do we use it? Clarifying this with more robust empirical data analysis is a key aim of this research project.

Summary

In this section I will present a brief summary of what I feel are important points in showing how the above findings have been reached. To do this, I will go through each chapter of the dissertation, as well as comment on one of the appendices. The structure of the dissertation is composed of six chapters.

1. An Introduction to 21st Century Complexity, Leadership, and Hard Stage Developmental Models
2. Two Psychometric Studies on Adult Development Scores and Hard Stage Requirements
 - a. Study 1A: Cognitive-Developmental Scores awarded by the Lectical Assessment System (LAS)
 - b. Study 1B: Ego development Scores Awarded by the Ego Development Scoring System
3. Additional Analyses Undertaken on Ego Development
 - a. Study 2A: Reanalysis of Earlier ego Development Data
 - b. Study 2B: Scoring Ego Development Scoring Exemplars with a Cognitive-Developmental Scoring System
 - c. Study 2C: Regression Analyses on the Cognitive-Developmental Scores and Ego Development Scores
 - d. Synthesized Discussion on Ego Development
4. The Complexity Gap Between Leader Reasoning Skills and the Task Demands of Their Roles
5. Longitudinal Growth in the Hierarchical Complexity of Leaders’ Reasoning Skills
6. General Discussion

There are also 15 appendices included. Most of these are technical supplements to the main work, however one merits a summary presentation, appendix A on Hard Stage Claims Made by Ego Developmentalists. As well, there are 30 pages of references, 58 tables, 42 figures and the total length of the dissertation is 457 pages.

Chapter 1

The introduction sets the stage or context for what is to come. The rationale for the study is clearly laid out. “Leaders in the 21st century are confronted with an unprecedented level of complexity” (p. 1) such as VUCA environments, so that “it is no surprise that complexity has been amongst the most significant challenges experienced by organizational leaders for over a decade” (p. 2). Three types of intersection between leadership and complexity are identified: complexity leadership theory, leader complexity theory and the focus of this study, adult development (AD) theory. Aiden lists numerous studies that ‘presume but don’t demonstrate’ the claim that higher ‘stages’ more adequately handle the complexity of workplace environments. Within AD, the chosen focus is on hierarchical complexity, given that many in the adult development community, theorists and practitioners, make either implicit or explicit claims that the stages they measure are hierarchically integrated and thus reflect an increase in the complexity of thought and or performance.

Aiden goes on to do a good job of describing the distinctions between cultural age-related models of development from soft and hard stage models. He then breaks down the theoretical and conceptual basis for each type of stage model and how they relate to the conception of complexity. From this, he identifies that hard stage models are most likely to show empirical evidence of the hierarchical complexity. This will of course have consequences for proponents of adult development, (myself included), who have made claims about complexity in leaders based on various soft stage models.

Aiden is precise in the claims he makes, being careful to indicate what his main concern is, that ego or other stage development theories “may not be well suited to sense-making about a particular type of complexity” (p. 25). Having laid out the type of complexity in focus in terms of the hierarchical complexity, this sets the stage for three research questions and a series of studies to ascertain answers to these questions.

- RQ1: Do cognitive developmental scores awarded by the Lectorial Assessment System (LAS), and ego development scores awarded by the ego development scoring system, satisfy hard stage requirements?
- RQ2: To what extent does the hierarchical complexity of leaders’ reasoning skills satisfy the task demands of their roles?
- RQ3: To what extent do leaders develop the hierarchical complexity of their reasoning skills during participation in various leader development programs?

Studies 1A & 1B and 2A, B&C all address RQ1. Study 3 addresses RQ2 and study 4 addresses RQ3.

Chapter 2

Chapter 2 is the core of this research project and will be examined in more detail than the later chapters. Its focus is on establishing how scores awarded by the Lectorial Assessment System and

the ego development scoring system, (primarily Loevinger's WUSCT), meet or do not meet the criteria for hard stage models. Study 1A focused on this question for analyzing patterns in scores provided by the Lectical Assessment System, and study 1B for patterns in scores provided by the ego development scoring system. The underlying aim is to establish construct validity in relation to addressing complexity.

Study 1A

Aiden starts by addressing the concern that skill theory / Lectica is too focused on a narrow conception of reasoning – i.e., ego development is more holistic. He cites several sources that show the actual holistic conception of cognition from Piaget on includes thinking, motivation, meaning, feeling and acting. The evolution of my own understanding of developmental psychology (apart from ego development theory), through the work of people like Kurt Fischer and Michael Mascolo, has made it clear that this field sees 'cognition' as much more than mere intellectual thinking or reasoning, but that it holistically integrates emotion, embodiment and dynamic environmental relations.

He goes on to describe several studies that note how Michael Common's Hierarchical Complexity Scoring System (HCSS) and earlier iterations of the Lectical Assessment System (LAS) meet hard stage requirements and have strong correlations with other instruments and scoring systems. Aiden proposes that it is worth seeing if a finer granularity of Lectical scoring can still meet hard stage requirements as operationalized through the unidimensional Rasch model. This arises from the observation (which my own experience agrees with) that using assessments at distinctions of one full stage of development are impractical, as they lead to suggested learning and skill development practices that can easily be outside the "Goldilocks zone" or Vygotsky's zone of proximal development. Being able to have trust in the validity of finer-grained scores, in this case quarter of a level Lectical 'phases,' enable the provision of more precise and usable developmental practice recommendations.

There is a detailed explanation of Rasch modeling, but I have to admit it is a bit outside my own Goldilocks zone. What I can say is that the description of the method used is very thorough. I did notice that I would have appreciated an explanation of the differences between figures 15 and 16 on page 83. As a non-statistically sophisticated person, I was not sure why the visually observable differences in the two graphs were not unpacked.

I appreciated that Aiden identified potential competing explanations that could account for the pattern of results obtained for Lectical Phase Scores, e.g., halo effect from scoring entire LDMA's at once. The second possibility mentioned relates to the narrow distribution of item difficulties, however this is something that could use future research to determine if relevant or not, given the different nature of the Lectical assessments. A related issue is raised related as to how the dilemma difficulties and prompts in Lectical assessments, in this case the LDMA (Lectical Decision Making Assessment) tend to fall in a narrow range, which could prompt respondents to respond within that similar range. A suggestion is made that "Lectica may consider revisiting its rationale for using items with similar difficulties" (p. 88). A fourth possible competing explanation put forward is that if a) Lectical Phases increased in elaboration rather than hierarchical complexity and b) there is a tendency for people to get the same score to all items, a similar pattern of results may be

obtained. As a result, it may be important for Lectical to rule out the possibility that Lectical Phases reflect an underlying increase in elaboration rather than hierarchical complexity.

The results of study 1A were not unexpected. Lectica has built a solid foundation for their assessments, and this shows in the move from measuring hierarchical complexity at the scale of full complexity levels to measuring the hierarchical complexity at the scale of Lectical Phases, or quarter levels. My own experience using LDMA's with clients reinforces that a phase makes a real difference in practice.

What I was a bit surprised about was the range of scores in focus in the study. While 10c² does represent what I have experienced as the low end of adult performances, at least for leaders, I was surprised by the distribution curve in Table 7. It showed the peak of the distribution curve to be in 11b. This might be due to the earlier LAS human scoring, which has been understood, in retrospect, to have fallen prey to a 'jargon' bias.³ Also that the data for this study came from assessments scored between 2006 and 2016, before CLAS (Lectica's Computerized Lectical Assessment System) was utilized. Since then, scores awarded in the level 11 range tend to be lower than when only human scoring was involved. The introduction of CLAS brought about a realization of how the use of constructs that were seen to occur after the density curve bottomed out, (see footnote 3), indicates non-assimilated or non-integrated terminology being used in many performances.

A second puzzle, possibly explained by this sample being focused on leaders and not the general population, arose for me from earlier information⁴ that indicated a distribution of Lectical scores among the general adult population that peaked in the 10c/d range, at 60%, while the 11a/b range containing only 30% with only 9% 11c/d and less than 1% 12a/b.

I do not find that this difference between scoring distributions of leaders versus the general adult population causes a major concern.⁵ The introduction of CLAS has helped understand scoring in a more objective manner and creates a cleaner or more objective look at what this core structure⁶ looks like in language.⁷ This move to lower and more narrowly distributed scores in

² Lectica's model uses a numerical labeling. What is important here is that level 10 represents abstract mappings, and level 11 systems of abstractions. As well, there are four phases, or quarter level substages within each major level. You can learn more about these levels [here](#).

³ You can read a bit about this [here](#).

⁴ I recall presentations from Lectica with these statistics and I have used them in public presentations. I dug around and found an internal document used by Lectica as part of supporting the training of consultants to use their assessments that had these statistics. It listed the relative frequency of 'zones,' or two Lectical phases among the adult population, presumably based on their data set at that time.

⁵ While searching for the above noted internal document I also came across one showing the same distribution curve Aiden uses for leaders.

⁶ You can find an article explaining this [here](#).

⁷ There is ongoing philosophical debate about the relationship between language and consciousness. What is important here is that Lectica, as well as any other such assessments, use language as a proxy for actual thought / skill structures. There are also conversations I have been involved in thinking about various limitations of this, such as how the use of metaphors might not stand in as a proxy in the same way, or how people using a 'construct-aware' consciousness might not be able to adequately express all of their cognition in language.

level 11 raises a number of questions that this study only lightly touches on and are some of the many topics for future research and discussion.

Study 1B

The first part of this study takes a deep dive into literature and studies relevant to understanding how the field of ego development (ED) has been studied over time. For instance, it is noted that ED does not use a Freudian conception of ego but is “conceived to be a synthetic process analogous to a psychologically oriented central processing unit” (p. 90). As such, Loevinger seems to provide more of “a descriptive framework than an explanatory scientific theory” (p. 91). This background helps provide context for examining how ED relates to the criteria for being a hard stage model.

First up are the criteria of unidimensionality, examined through a lot of references to studies that show that ED appears to be more multi-dimensional. Aiden reviews the main players in the ego development theory lineage’s logics. He notes that Loevinger did not come up with an underpinning logic. He uses Torbert’s conceptions of stages and various descriptions of the logic behind them, such as ‘norms rule needs’ evolving in the next state to ‘craft logic rules norms,’ as one example of how the theoretical underpinnings of ego development are constructed. He identifies that these may also explain other processes related to more conventional, Freudian ego processes like sublimation. I found it interesting to note these distinctions between the ED conception of ego and Freud’s.

Aiden also draws on Susann Cook-Greuter’s notion of changes in perspectives on the self as a logic.⁸ He notes that she “contends that her logic [moves] ego development theory from a ‘soft’ stage theory to one more akin to a ‘hard’ stage type according to Kohlberg’s distinction (Cook-Greuter, 1999, p. 52)” p. 93). This is one of the core claims being addressed in this dissertation. The move to produce models and assessments that meet this hard stage criteria, for whatever variety of reasons, are being contested and examined here.

Aiden claims that: “The ego development scale is contended to be a unidimensional ordinal scale consisting of nine hierarchically integrated stages which satisfy hard stage requirements” (p. 94). The implication of the term ‘contended’ seems to me a bit different than the above quoted phrasing of ‘more akin.’ It is not clear where is this clear statement of claiming ED as a hard stage comes in. While I noted this slight difference of connotation or implication, I later saw that Aiden addresses this at length in appendix A.

There is a discussion of how ED theory and its measures are interwoven, developed together through an iterative bootstrapping, and that issues with the measure might have implications for the theory as well. The scoring system is described, and an elaborated set of constructs proposed to be measured are described.⁹

⁸ You can find the detailed descriptions in table 14 on pages 96 and 97, along with a note at the bottom elaborating on the sources used.

⁹ Tobert’s 2013 article *Listening into the Dark* provides on of the more recent and thorough efforts to show validity and efficacy for his developmental action inquiry version of ego development. Aiden reviews this later, on pages 115 through 117 of his dissertation.

Aiden moves on in his evaluation of hard stage claims, noting that ego developmentalists “consistently espouse the hard stage properties of their stages as outlined in Appendix A, but their evidence is equivocal” (p. 100). He then brings more of his findings to bear on the claim of unidimensionality, noting that only 20% of the variability in SCT scores is attributable to the construct of ego development itself. I saw that this comes from Loevinger and Wessler (1970) themselves and various other empirical studies!

It was fascinating to read about early contemporary studies of Loevinger’s work. There were more studies that also corroborated the finding that ego development only seems to account for approximately 20% of variability in scores awarded to SCT stems. The study of early and late adolescents introduced secondary factor of word count. Later studies showed a possible higher order construct beyond ego development itself that was strongly related to interpersonal style. I found that this made intuitive sense to me.

The hard stage criterion of invariant sequence is seen to be supported by a variety of earlier research done covering cross-sectional, longitudinal, interventional and cross-cultural evidence. This ends up being the only hard stage criterion that is adequately met.

The examination of the hard stage criteria for structure of the whole looked at variance in individual stem scores and how they tend to be washed out in final TPR aggregated scoring. I know this from my own report on taking a sentence completion test (SCT), where I responded in a variety of ways and had scores across almost the full spectrum of stages. Loevinger’s own admission of structured wholeness being a ‘working assumption’ is noted as well as a lack of subsequent research attempting to validate this. Three potential reasons are explored; measurement error (not seen as adequate), decalage (to which the TPR single score stands in contrast, raising the question of why people are not provided with a more varied result), and growth over time, where earlier reasoning is still able to be accessed (somewhat plausible).

The section examining the criterion of stages being qualitatively distinct opens with doubts on the existence of distinct stages from Loevinger herself later in her career. It goes on to point out that published reliability scores would indicate the ability to distinguish between 3 to 4 statistically distinct stages, not nine even though the raw scores awarded to SCT stems may suggest otherwise.

The hard stage criterion for hierarchical integrations is also found wanting. First, the inability to ‘fake’ higher scores (shown by a number of studies) while interesting, does not make it clear that the stages represent a succession of hierarchical integrations. A study looking at the correlation between SCT scores and cognitive developmental scores showed a much lower correlation, ranging from .35 - .52, far less than the .86 - .92 correlation found between scores based on hard stage models. Further discussion reveals a range of other considerations that make it less plausible for ego development stages to meet this hard stage criterion.

Aiden provides justification for using the same criteria in study 1B as in study 1A, by noting that ego developmentalists claim that hard stage criteria are met. (Again, see comments on appendix A for more on these claims). He presents a series of justifications for how he analyses data to address each of these criteria. There is also a thorough description of how participants and data was gathered and prepared for analysis. There is a careful, step wise set of procedures

described to make sure the data being analyzed was appropriate for this process and several steps in this process are described.

Step 3 was the big one, a direct test of hard stage requirements. Lots of data and numbers are presented in table 25, but us non statistically fluent, this could use more immediate explaining to understand. Some more direct interpretation comes eventually where the “analyses suggest that some subsamples display a moderate amount of multidimensionality. This suggests that the SCT is an impure measure of ego development ...” (p. 143).

The data cleaning procedures undertaken in Step 1c resulted in the need to delete some items and stages to enable to the whole data set to be analyzed i.e., from 55 to 42 stems and from 9 to 7 stages.

More analysis led to finding that there was not sufficient ‘white space’ between item thresholds to identify qualitatively distinct stages on the Rasch logit scale. Furthermore, figures indicated that there was not sufficient reliability to measure seven statistically distinct stages. Reliabilities indicated that as few as three statistically distinct stages may be measured for some subsamples. Four to five stages being able to be measured was the end verdict of the overall data analysis.

Clear visuals figures show evidence that these stages do not appear consistent with the criterion of hierarchical integrations. These figures are in stark contrast to the ones shown earlier for the Lectical Phase Scores. Spurts and plateaus of development were not observed in the analysis of scores awarded by the ego development scoring system. In the end, only the criterion for invariant sequence was satisfied, not the criteria for unidimensionality, qualitative distinctness, structured wholeness or hierarchal integrations.

Aiden explores a few explanations for these results and shows it to be “an artefact of how the hard stage requirements were operationalized and the Rasch procedures employed” (p. 151). Then an explanation that SCTs / ED are soft stages is proposed, which fits with Kohlberg’s analysis and discussions with Loevinger.¹⁰ The question arises as to why ED theorists make the claim for meeting hard stage criteria. This is explored more in the synthesized discussion on ego development. For myself, I would be curious if these people would today make these same claims with the same emphasis.

A key issue discussed is the focus on ego development as a unified construct and how this affects how it is being measured. ED theorists appear to, over time, make increasingly broad claims about how much ED actually encompasses as a holistic construct. This would seem to add the possibility of more fuzziness, and potentially show up in the analysis as multi-dimensionality.

Aiden raises questions about the single stage TPR final score awarded, given the noise and multi-dimensionality in the results. The issue of ‘decalage’ or variability of performance in different domains (not to mention under many different circumstances), could be pointing to some of the underlying differences between the theoretical constructs in the theories. This is most

¹⁰ See chapter 3, especially pages 236-249 in Kohlberg, L. (1984). *The psychology of moral development: The nature and validity of moral stages*. Harper & Row.

exemplified to me in the difference between the conception of a ‘center of gravity’ in ego development and the core principle of ‘variability of performance’ in dynamic skill theory.

While the discussion points Aiden raises are clearly focused in relation to the purpose of the study, I find myself wanting to broaden out the discussion and take some of these points up after completing this summary of the dissertation.

Chapter 3

This chapter undertakes a new set of studies aimed to better explain and understand the results of study 1b.

Study 2A

Study 2A revisits Cook-Greuter’s 1999 dissertation data set, to re-analyze the data with more modern statistical means. This study, along with Studies 2B and 2C which are discussed below, provided a way of ruling out the possibility that the findings of Study 1B were an artefact of the Rasch procedures used there. From this and due to the data in the Cook-Greuter’s dissertation being aggregated, the same Rasch method used in study 1B was not appropriate. This meant that only invariant sequence, qualitatively distinctness and structured wholeness could be directly tested, but not unidimensionality and hierarchical integration. However, Aiden did analyze the validity and soundness of Cook-Greuter’s argument in support of hierarchical integrations between stages, given that statistical analyses could not be performed.

Figure 29 provides a simple boxplot of the median scores of Cook-Greuter’s data, and these clearly show significant overlap between achiever and individualist scores, as well as the construct aware and unitive ones being almost identical, which would call into question the overarching conclusion drawn in Cook-Greuter’s dissertation, which was to show a distinction between stages at the higher levels.

Invariant sequence analysis showed that there is a statistically significant difference between the average scores for successive ego development stages (see pages 161 / 162). Here, it is shown that the expert and achiever mean scores fell into the Expert stage, with a difference of 0.59, with similar results for the individualist and strategist falling into the Achiever stage with a difference of 0.49 of a stage. Another statistical test was done which did show significant differences between the distributions for successive ED stages, yet the visual overlap from earlier figures raises serious questions still to be addressed. Yet another statistical test shows between 74 and 84% overlap between adjacent stages.

The discussion of this analysis concludes by noting that once again, only the criteria for invariant sequence was satisfied, and a variety of potential reasons are explored for the gap between the results and interpretations in study 2A with those of Cook-Greuter in her 1999 dissertation.

Study 2B

Study 2B describes claims of mapping between models from Cook-Greuter and Torbert, namely to Kohlberg's stages of moral development and Model of Hierarchical Complexity (MHC), as well as Lectical levels. An analysis of Lectical Phase Scores awarded to exemplars from Loevinger's scoring manual did not show direct correspondence between ED and Lectical scores. The meant that these ego development exemplars did not show a significant increase in hierarchical complexity between successive ego development stages.

The discussion of study 2B indicates that the stages from expert through unitive all have a mean Lectical score within level 10, equivalent to formal operational thinking. Aiden notes that his "findings suggest that Construct-aware and Unitive exemplars may be poetic expressions of formal operations which fail to fulfil the requirements of higher order hierarchical integrations" (p. 182). He goes on to note that Cook-Greuter (1999, p. 96) admitted that her conceptions of these later post-autonomous stages,

"can readily be proved false if there is evidence that individuals at earlier ego stages are capable of making the kinds of distinctions at the required level of complexity that [she is] suggesting, are reserved for the most advanced meaning makers." Findings suggest that not only have her notions about postautonomous development been 'proven false,' but the general claim that stages satisfy the requirement of hierarchical integrations has also been challenged. (Thornton, 2023, p. 182)

One of my own observations is that 'thinking' in general, and linguistic proxies for thinking, (not to mention the ongoing debates about the relationship between consciousness and language – does language structure our consciousness / thinking, or does it channel or represent it?), is to point to the limitations of all language-based attempts to measure subjective constructs. So, from this, could it be that while the Lectical assessment system does a good job of measuring hierarchical complexity of linguistic representations of thinking, by being precise and specific, ED attempts to get at more of the non-linguistically represented aspects of consciousness, and from this suffers when measured against stricter constructs? Of course, ego development assessments also have to use linguistic representations of thinking as a proxy for the constructs they are hypothesizing to measure.

The evidence accumulated in these studies may well challenge claims made by ED theorists and practitioners that were made earlier. How they respond to the findings of this dissertation will be of interest. At the same time, I hope it can open up a more informed conversation to better scope out the territories being assessed by all such measures and inform practitioners in a way that can alleviate very real concerns of overly eager and simplistic claims being made in the coaching and consulting world.

Study 2C

This study aimed to address an ongoing challenge in making comparisons between measures by scoring responses to SCT protocols with the ego development scoring system, CLAS, and other theoretically related approaches. It also attempted to identify which of four key variables could

account for the greatest amount of variability in SCT scores i.e., age, number of perspectives taken, unique word count, or hierarchical complexity. The premise being that if SCT scores meet hard stage requirements (and therefore reflect the underlying construct of hierarchical complexity), then the other measures such as age, word count and number of perspectives should not account for a significant amount of variability in SCT scores beyond that accounted for by CLAS scores. A very thorough set of procedures was used for CLAS to be able to address more ‘poetic’ later stage responses and identify variables such as unique perspectives and non-repetitive total word count. In my view, this study presents some of the most important and impactful findings presented in the context of Aiden’s dissertation.

The data preparation and identification of factors to use in the regression analysis was clear and useful. The procedure started with the CLAS scores and sequentially added in variables to determine their predictive value. SCT scores shared 27% of variance with CLAS scores, unique perspectives added an additional 1% of shared variance and unique word count added an additional 2% of shared variance. Considered individually, SCT scores was the strongest predictor of CLAS scores with a standardized beta-coefficient of .35, compared to standardized betas of .27 for unique word count and .02 for unique perspective taken.

The regression analysis was then run the other way, starting with the SCT scores, and looking at the predictive value of CLAS scores, unique perspectives and unique word count. CLAS scores shared 27% of variance with SCT scores, but when adding in unique perspectives the shared variance rose significantly to 42%. When unique word count was also added to the model, a further 11% of shared variance was added so that CLAS scores, unique perspectives, and unique word count jointly accounted for 53% of variability in SCT scores. Considered individually, unique word count was the strongest predictor of SCT scores with a standardized beta-coefficient of .53, compared to standardized betas of .23 for CLAS scores and .08 for unique perspectives. As a result, unique word count was found to have twice the predictive power compared to CLAS scores.

What is clear from these results is that while SCT scores are somewhat related to hierarchical complexity as measured by CLAS (i.e., 27% shared variance), SCT scores appear to be more significantly related to unique perspectives and unique word count, thereby suggesting that hierarchical complexity is not the primary construct being measured by the SCT.

The discussion that follows reinforces the findings reported in studies 1B, 2A, 2B – all of which suggest that the SCT scores do not appear to consistently meet hard stage requirements. Hierarchical complexity only accounts for between 12-27% of the variability in the SCT scores, while perspectives and word count added a further 26%.

The synthesized discussion of these findings describes how four different studies, employing different samples and research methods all came to a similar conclusion, that SCT / ego development scores seem to “operationalize approximately four or five cumulative stages that are moderately related to hierarchical complexity and more strongly related to the number of perspectives and/or words used by test-takers” (p. 204). The question then arises, why has this not been clear before?

One possible reason Aiden explores is that of all the previous studies examined, Cook-Greuter's 1999 dissertation was one of the few that attempted to explicitly test hard stage requirements, and "her conclusions depended on visual analysis of SCT distributions and what appeared to be a logical fallacy. Her conclusions were not supported when her data were reanalyzed using quantitative procedures" (p. 204). A second reason proposed is that hard stage requirements were over generalized to psychological constructs to which they might not apply, e.g., the ego or worldview. Aiden argues that some of the psychological constructs that are included in considerations about ego development (e.g., feelings, defense mechanisms, personality structures, etc.) have not necessarily been shown to be subject to hierarchical integrations and therefore, cannot necessarily be combined to yield a single variable. As well, the attempt to include so many variables as are included in the ego construct leads to an inevitable contradiction with trying to have a sound hard stage criteria meeting outcome. "Findings from this thesis suggest that it results in an increasingly integrative conception of the self that is decreasingly related to hierarchical complexity" (p. 205). Even Loevinger recognized that was an open question as to whether ego development is made up of several unified strands of development or if those strands of development needed to be considered and measured separately.

A third reason proposed is the general acceptance of statements by Loevinger proposing that the various accounts of development are alike enough to suggest a common component. From this, others tend to build on the sentiment and organize further arguments to reinforce this view. Finally, the possibility of a simple conflation between general sequences and hierarchical integrations is addressed. SCT scores do display an invariant sequence, but that does not imply that ego development stages form a sequence of hierarchical integrations. An acknowledgement is made that SCT scores are related to various leadership phenomenon, but this may not be attributable to the SCT's ability to measure (or not measure) hierarchical complexity.

Several potential rebuttals are addressed, such as:

- Ego development going beyond the mechanics of cognition, (seen as a limited understanding of the Neo-Piagetian conception of cognition).
- A focus on reasoning (similar rebuttal).
- The narrower construct in focus in cognitive-developmental assessments making it easier to display stronger psychometric properties, (examples of diverse psychometrics are provided, showing that what determines good psychometrics are not limited to these concerns).
- That cognitive development might present barriers to leadership studies, (this appears to be an appeal to emotion. However, the quote used for this, in my view, is referring to a carefully worded acknowledgement of types of resistance that for me, have more to do with poor contextual awareness and a lack of appropriate framing of the work).
- That because ego development includes cognitive development and more, not all these aspects can be expected to correspond with cognitive development, (multiple rebuttals,

including that these aspects would need to show increases in hierarchical complexity, which they don't). And finally,

- That people cannot grasp concepts transcending their own level of ego maturity (shown as a presuppositional fallacy calling into question the findings described in this thesis and potentially an ad hominem fallacy that is sometimes directed towards scholars who critique ego development theory and associated forms of measurement).

I want to take a moment here to acknowledge my appreciation of the thoroughness of Aiden's explorations and considerations that he addresses. I have, over many years, gone through adopting many of the perspectives critiqued here. Having them carefully considered, explored, analyzed with robust tools and procedures and presented with clear deliberations has helped me mature and clarify my own understanding of many issues in the field.

Chapter 4

Having addressed the initial set of concerns related to claims for how Lactical and ego development measures relate to hierarchical complexity, Aiden now turns to the second research question, which addresses the issue of the complexity gap between leaders' reasoning skills and the task demands of their roles.

The concept of the complexity gap is illustrated by using the water conservation experiment. Then a set of arguments from philosophical, management and leadership and developmental psychology literature are reviewed. Aiden looks at perceived, inferred and conceptual capability gaps. Notable are references to Elliot Jacques' requisite organization work and Kegan's 'In Over Our Heads' concept, along with Dawson and Stein's research on US government managers.

Study 3 used a set of Lactical assessments where there was also sufficient data to give an indication of the complexity of role of the test takers. A table described the skills required and range of Lactical scores determined as necessary for performing these skills, with distinctions between executive, senior, upper and mid-level leaders.

The findings, (with a similarly robust set of methods and procedures described), show that mid and upper-level leaders appear to operate with adequate levels of cognitive complexity for their roles, while there is an increasing gap for senior and executive leaders. Several factors are explored that may account for this, including the sample itself, biological factors related to neurological functioning, organizational factors such as time in role or leader development opportunities and finally organizational culture factors, where constraints may be in place that do not reward more complex reasoning.

The conclusion presented, that "it seems reasonable to suggest the leaders in the 21st century are experiencing a complexity gap" (p. 246), while apparent, misses explicitly mentioning the distinctions noted above between mid and upper level managers and those in senior and executive leader positions. While this study does give an indication of where a complexity gap is most likely to be found, it appears to me at least that there is much more work to be done in this area to actually

understand this phenomenon and begin to understand how to address it. This last issue is then the focus of the final study in chapter 5.

Chapter 5

Chapter five examines study 4, about longitudinal growth in the hierarchical complexity of leaders' reasoning skills. It takes on the thorny and challenging question of development over time – how to not only do it but measure it appropriately. Given the case made thus far for complexity of reasoning, it is as expected, that this study also draws on existing data sets from certified Lectical consultants' work.

A survey of the literature and conceptualizations surrounding this topic is undertaken to help narrow the focus and identify an appropriate target for this study. Part of this literature indicates that while growth in the complexity of reasoning is possible, it tends to occur more often and faster earlier in life. The description of studies of growth over time due to specific interventions is a nice overview of a range of efforts that have been undertaken in this area. However, this overview reveals the limitations of knowledge in this area, due to low sample sizes, lack of control groups and mixed methodological designs and interventions. This sets the stage for study four to try and make progress on generating a more robust and well supported understanding of this question.

In describing the details of the several statistical analyses performed,¹¹ I noted the finding that for each level of development gained at test-time 1, the average rate of growth per month occurred more slowly. You can see this in Lectica' growth curves, derived from the vast data set Lectica has available, including longitudinal data.¹² One result noted was that women grew faster than men! No real surprise there.

There are some interesting findings along the way in this chapter. The number of assessments taken positively impacted and to a degree predicted the average rate of growth, while the time spent on instruction received did not. Formative use of assessments also had a significant ability to predict the average rate of growth per month. An analysis of two programs that had enough participants revealed significant average growth rates. After this, a more exploratory analysis was undertaken with statistical constraints used to enable the inclusion of smaller numbers from the other programs.

An average rate of growth of 0.07 Lectical level per year was found. This rate of growth was significantly moderated by some pedagogical characteristics, such as the number of assessments used, the formative use of such assessments, but not time of instruction. The discussion proposes that findings indicate even though direct causal attributions could not be made “that the process of reflective abstraction may have been activated” (p. 279) and that “development programs may have played a role in leaders' growth” (p. 281).

¹¹ I wanted to make a note here regarding the rationale for Lectica making LAS and CLAS scores equivalent. I believe this could be questioned, given the lowering of scores in the upper range since CLAS came online. While this study was done early in the introduction of CLAS, it is worth noting there may be implications from what has been learned more recently in this area that could have an impact on the findings here.

¹² You can read more about this [here](#).

Chapter 6

In chapter six Aiden presents a general discussion of all the findings. The punchline is as noted in my introduction, that “ego development scores may be a stronger reflection of a soft stage model which increases cumulatively in the number of perspectives taken or the number of unique words used by test-takers” (p. 283). The growth of complexity is seen until getting to executive leaders. The complexity gap was noted to be primarily related to senior and executive leaders.

The bootstrapping development of ED between theory and measures can be seen to have contributed to some assumptions being propagated over time, which may need revisiting now considering this current analysis. Near the end, Aiden presents the implication that the findings of the studies he conducted have for not only measurement, but also potentially theory of ED or at least its relation to cognitive developmental theory.

Aiden makes some strong claims such as, “the findings reported on measurement of ego development ... may imply that the central tenets of ego development stages may need to be revisited” (p. 289) and “ego development scores do not seem to be well suited to making inferences about leaders’ ability to navigate complexity” (p. 289). These are serious statements that challenge a reasonable body of research, saying that some of the fundamental theoretical and measurement assumptions and assumed linkages to cognitive complexity, and thus by proxy, a key element of leader capability. There are also important implications for the variability of scoring of ED stages in relation to different stem types or domains; “ego development stages have different meanings in different contexts” (p. 290).

Aiden presents some interesting speculation about explanations for why “ego development scores predict some complexity related leadership outcomes” (p. 290). The prevalence of randomness itself is explored as an explanatory factor in achieving outcomes. For me, this simply points to how there are generally deeper causal factors about things in life that we might make sense of in terms of being complex, chaotic, or random, given the limits of our conceptual and perceptual lenses.

Aiden also puts forward explanations of how a broader set of perspectives being held (supported by the soft stage model) could increase a leader’s capacity to lead in complex conditions. There is also a nod to how the presence of leaders, characterized by things like warmth, awareness and sagely interpersonal styles, could contribute to conditions under which performance improves. (This seems quite plausible to me).

He notes that “perspective taking may have less in common with hierarchical complexity than what has typically been assumed in the adult development literature” (p. 291). He also indicates the implication that ego development may not constitute vertical development in the manner that the term has been used in popular literature. (Vertical development has become a common phrasing being used to indicate developmental growth among several practitioners. I have my own reservations about the oversimplifications involved in how people use such a construct and tend to avoid using it).

Ego development as a “perspectival theory” (p. 291) is described with the caveat that:

This is not to suggest that ego development is inferior to cognitive development as a general approach to adult development. It is to suggest, however, that ego development stages may not exhibit a particular property and they may be ill-suited to applications related to complexity in particular. (p. 292)

This contextualization and delimitation of EDT and assessments appears to me to balance the recognition of value in these models and measures as well as supporting a more refined and precise utilization of them in practice. (Various conversations I have had with ED practitioners gives me an impression that some are already moving in this direction and appreciate the support for this that Aiden's research is providing).

Several future implications and possibilities are described for the support of cognitive developmental hard stage measures, including the importance of less than full stage granularity being able to be operationalized. A claim is made regarding study 4 that leader development “programs were more likely to result in growth because they incorporated learning material that is aimed at a higher level of hierarchical complexity” (p. 295). My question is in relation to the claim of “because they incorporated learning material ...” – if this is referring to Lectical or other assessments, then this should be spelled out. As it reads, it is inferred, or a claim is being made in general about the content of the programs that were evaluated in this study. Reflective abstraction is focused on in this discussion and while Lectical assessments might be a good structure for encouraging this activity, ego development assessments, used well by practitioners, might also foster this. I would have appreciated this being made explicit.

Aiden makes several strongly supported claims related to the practice of using ED measures for addressing complexity. Considerations are raised related to inaccurate information being provided to clients, based on the disparity between claims related to ED models and the findings of this thesis and also the historical literature that was reviewed in the context of Study 1B. As well, this information, about claims of ED theory and its implications, could misinform clients and potentially contribute to adverse impacts on self-concept and leader identity. There is even a suggestion that a revised version of ED theory and SCTs could be used to foster skills like interpersonal maturity.

Implications for practice of addressing the complexity gap are also discussed, including recruitment, promotion, succession etc. A rich set of recommendations of practices for leaders are also described, including systemic and collective cultural leadership practices. Also, “it may be possible to design practices which integrate behavioral change with the development of hierarchical complexity” (p. 298). My own experimentation with helping leaders design such practices for themselves, based on feedback from Lectical assessments, indicates to me that this could indeed be a powerful form of intervention.¹³

Another implication Aiden explores is that given the rate of development of cognitive complexity found, addressing the complexity gap will require finding people who have less of a personal complexity gap, or are given more time to develop, as well as developmentally designed interventions to support such growth. As well, leader development likely needs to be supplemented

¹³ You can read more about one such experiment [here](#).

by collective practices such as collaborative decision making. Overall, providing opportunities for meaningful reflection is key to developing more adequate capacities for navigating complexity.

Near the end of this general discussion Aiden presents a set of broader implications for the fields of adult development and leadership. This includes a list of big assumptions (BA) in adult development theory and leadership.

BA1: Adult development stages reflect and underlying construct of complexity.

BA2: Stages from one theory directly correspond to stages from another theory.

BA3: Choosing between one theory and another is a matter of personal preference.

BA4: Adult development assessments are directly measuring complexity of constructs related to complexity.

BA 2 is seen in many tables of correspondence between models, lining stages up. Overall, these BAs are seen to stand on questionable intellectual ground.

Then Aiden presents a very nice proposition for using a structured process of differentiation and integration to contribute to a process for growing the field of leadership development in general. Five transition steps are described, adapted from the MHC and that align with Basseches and Mascolo's TACS (thesis, antithesis, conflict, and synthesis) model:

- Thesis – the starting point, e.g., theory x is best suited to make sense of leadership.
- Deconstruction of the thesis – seeing the limitations of theory x to make sense of complexity related phenomena.
- Antithesis – theory y might be the answer!
- Relativism – contextual fit, or theory x sometimes, theory y other times.
- Smash – coordinating theories x and y enables a larger set of outcomes to be accounted for.
- Hierarchical integration – we make sense of adult development as an emergent property through the coordination of different theories with respect to both structure and content.

Using this process to help evolve how we come to terms with the range of distinctions and domain limitations and fit of various theories relating adult development to leadership can help establish a more robust and appropriately integrative understanding. By making clearer and cleaner distinctions about what each model or theory or assessment can do, we can use them in ways that support the needs of leaders. Aiden invites this kind of process to create a more “intellectually spacious conversation” (p. 303) for the field as a whole.

Aiden's reflections on how to make good use of these two theories includes the possibility of regarding them as orthogonal rather than isomorphic, that later stages of ego development might

reflect a highly deconstructed ego, but that this does not automatically imply high levels of hierarchical complexity.

Another consideration is the very practical need to work on the granularity of measures, as since development is in small steps, being able to target those appropriately is important.

Aiden presents some further reflections on the need to include but also go beyond hierarchical complexity and include a robust set of considerations to account for how leaders can navigate complexity.

A final part of the general discussion includes considerations about the four studies. Aiden does note updates about CLAS scoring and potential implications for the LAS. (See footnote 8 as well). It would be useful to replicate study 1A with the Lectica's revised assessments systems to see if the same findings emerge. Potential means of improving ego development scoring are revisited briefly as well. The possibility of applying the methods used in studies 1 and 2 to a broader range of adult development assessments (e.g., Kegan's, Stages, Graves) is briefly explored. Applying the same methods of analysis to these would be interesting and further enable robustness in the intellectually spacious conversation.

Aiden notes that in relation to studies 3 and 4, one unanswered question is how large a complexity gap will impact leadership effectiveness? What role does social scaffolding play in this, or other mediating considerations? As well, the type of data used to analyze management level complexity may not apply in matrix organizations or other contexts, and there might also be other unaccounted for mediating factors. Thus, the weaknesses of this study are acknowledged, while the findings are still of clear interest.

In a similar manner, there are several considerations unable to be analyzed or taken into account in trying to determine how much development happens from leader development programs. A limited sample, lack of details about methods employed etc.

Appendix A

Appendix A takes a deeper look at hard stage claims made by ego developmentalists. There are a set of extensive quotes from Loevinger about stages, criteria etc. Aiden then makes a strongly based statement, "Loevinger asserted that there was a direct one-to-one correspondence between her stages, Piaget's stages of moral reasoning (Loevinger, 1976, pp. 79-85) and Kohlberg's stages of moral reasoning (Loevinger, 1976, pp. 118-122)" (p. 354). Yet just above, in a quote from Loevinger, she uses the phrase "The stages of the two conceptions can be set in approximate correspondence" (p. 354). Approximate is not the same as one-to-one. There may be other interpretations available, but my read of these excerpts gives me the impression more that Loevinger was exploring how Piaget's criteria appeared to inform the evidence of stages she was discovering through her bootstrapping method.

Similar descriptive statements are extracted from Torbert's work with a similar one-to-one correspondence claim noted. Again, my read of those statements would not lead me to make such a strong assertion. It might be that the citations made that follow this assertion point to direct quotes

claiming one-to-one correspondence, but the quotes taken above do not, in my view, provide sufficiently clear evidence of this.

On the other hand, the first direct quote from Susann Cook-Greuter is very clear in asserting that the “evolving perspective on the self, as measured by the SCT, can be shown to satisfy the demands for hard-stage theory” (Cook-Greuter, 1994, p. 121, in Thornton, 2023, p. 348). Further clear and direct claims are quoted with specific use of terms like ‘stages hierarchically integrating’.

This appendix does provide a larger set of reference points to show that there are statements in various pieces of literature making what can be interpreted as hard stage claims for ego development measure. Given that most of these statements are from publications 20 - 40 or more years ago, my impression today is that I suspect such claims would not be held to as strongly. Given the evidence provided by Aiden in this research, I hope that a robust conversation among theorists and practitioners can explore how to make best use of the findings and implications put forth here.

Summary Reflections

There are many layers of reflections going on for me as I come to make some brief summary reflections on this piece of work and its implications. A broad observation is that this is a significant part of what might be called a second-generation contribution to the field of adult development and leadership. As the early pioneers in the field move towards handing over the work to others, various voices are emerging who are taking the field in at times new directions and sometimes, as in Aiden’s case with his dissertation, (he is also taking the field in new direction, but that is another matter for another time), stopping to help us revisit long-held assumptions.

Aiden’s work in this research project focused on clarifying what claims and distinctions can be made with two models and measures in the field, ego development through sentence completion tests and Lectical assessments using dynamic skill theory and hierarchical complexity. It helps us refine how we use the term complexity, challenging us to bring rigor to our use of it and more importantly, to carefully consider how we apply any of these tools in practice. This focus is appreciated, and my hope is that it will inform ongoing conversations among both theorists and those who train practitioners to use such tools.

Aside from some of the small comments I made along the way in the above summary, I also notice other questions arising. What hasn’t been addressed that might also be important to take into consideration in how we view working with leadership and the development of capacity to handle complexity? Aiden does indicate a few points along the way relevant to this, however the scope of a dissertation generally means that it is not possible to adequately address everything.

Two aspects come to mind for me. One is the role of emotions in development and capacity in general. Aiden does point to the holistic nature of Piaget and neo-Piagetians conception of skills and cognition, including emotion. However, I am seeing more and more work that would place the role of emotions more centrally in the development conversation. How might the questions Aiden examined look if it was also possible to include robust ways to understand and include how emotions affect performance?

This leads into my second consideration, that of context. Again, Aiden does point to this in passing. However, I also am more and more aware that performance is mediated by contextual factors. Assessing individuals in isolation from the dynamics of social relationships, organizational culture and systems and so on can lead to an over emphasis on the individual and miss addressing the ‘knowing – doing gap.’ Just because someone can ‘perform’ on a given assessment does not mean they will always be able to utilize their optimal capacity. I do believe that such individual assessments of capacity, whether in the form of a soft-stage ego development, or action logic assessment or a hard stage, e.g., Lectical assessment, can give a very good sense of the range of performance ‘in the wild’ that can be anticipated. Thus, it is important to address the questions Aiden has gone into.

At the same time, is the distinction between hard and soft stage criteria the most important consideration? Given the two considerations described above, (and more that could be thought of), might there also be more important question to be addressed? I’m reminded of a question I once heard: Is it more important to be right or in right relationship?

Both are clearly important, and my belief is that the latter enables the former to be held in a manner that allows for healthy inquiry and curiosity. Aiden’s work is likely to stir up serious conversations in the adult development and leadership community. It is my hope that we hold each other well in this process.