Evaluating Mindset as a Means of Measuring Personal Innovativeness

by

Dennis A. Stauffer

reprinted from

International Journal of
Innovation Science

Volume 7 · Number 4 · December 2015

Multi-Science Publishing
1757-2223

2) *These two mindsets are not mutually exclusive but it's possible to distinguish between them. Each of these mindsets prompts a variety of beliefs, values and behaviors that are logically connected to one mindset or the other, such that these manifestations can be used to detect and measure the degree to which a person holds one mindset or the other.*

1.2 A Fundamental Distinction

In previous writings dealing with the human implications of this theory [9], I have referred to the Status Quo Cycle as a *Knowledge Loop* because it relies on existing knowledge. I have referred to the Innovation Cycle as an *Insight Loop* because it produces fresh insights. I have called following the Innovation Cycle or Insight Loop *Clockwise Thinking*, and following the Status Quo Cycle or Knowledge Loop *Counter Clockwise Thinking*. This is based on the way I have mapped these patterns (Fig. 1), but it also alludes to the fact that these two patterns flow in essentially opposing directions. The Status Quo Cycle is propelled by a sequence of prior causes, whereas the Innovation Cycle is propelled by the outcomes it produces. This directionality is not just metaphorical; it is supported by this research, as I will explain.

The dichotomy between these two cycles can be summed up in human terms as the degree to which someone is willing and able to systematically revise their own thinking and behavior. This is more than just learning in the educational sense, which often entails the adoption of a set of beliefs, accepting known facts or becoming competent at using various tools. Rather, it is about revising one's views based on new theory and evidence and investigation…recognizing new patterns, developing new theories and making discoveries. It's about the ability to learn new things for the first time—ever.

A number of noted authorities have described comparable dichotomies in a variety of contexts, apparently reflecting a similar fundamental insight about human cognition.

- Philosopher Alfred North Whitehead [10] distinguished between a *static* and a *dynamic* world view, in terms of both the assumptions one makes about the nature of reality, and how one comes to know that reality.
- Gestalt Psychologist Max Wertheimer [11] distinguished between *reproductive thinking*, the solving of problems by using prior knowledge, and *productive thinking*, the solving of problems with new insight.
- Educational psychologist Jean Piaget [12] distinguished between *assimilation* and *accommodation* in how students learn, *assimilation* being taking new data and incorporating it into a student's existing schema and *accommodation* being the adjusting of a student's schema to fit the data.
- Former head of the American Psychological Association, Raymond Cattell [13] distinguished between *crystallized intelligence*, the ability to use skills, knowledge and experience and *fluid intelligence*, the ability to recognize patterns, draw inferences and discover new relationships.
- Business and education theorists Chris Argyris and Donald Schön [14] expanded on earlier insights by Bateson [15] and Ashby [16] when they distinguished between *single loop* and *double loop learning, Single loop learning* is problem solving and taking action within predetermined assumptions and constraints. *Double loop learning* is problem solving and taking action by challenging those assumptions and constraints.
- Nobel laureate Daniel Kahneman [17] is one of many psychologists who have described a duality between *System 1* (or Type 1) cognition, which is predominantly subconscious, automatic, heuristic and habitual, and *System 2* (Type 2) cognition which is more effortful, reflective and deliberative [18].

These pairings are not making exactly the same distinctions and yet they all appear to depict a similar underlying polarity in how we think and behave, one that parallels the respective characteristics of the Status Quo and Innovation Cycles. Indeed, these pairings appear to be manifestations of these cycles.

Table 1 Parallel Concepts

Source	Distinction	
Stauffer	Status Quo Cycle	Innovation Cycle
Whitehead	Static World View	Dynamic World View
Wertheimer	Reproductive Thinking	Productive Thinking
Piaget	Assimilation	Accommodation
Cattell	Crystallized Intelligence	Fluid Intelligence
Argyris and Schön	Single Loop Learning	Double Loop Learning
Kahneman et al	System/Type 1	System/Type 2
Stauffer	Linear	Iterative
Stauffer	Knowledge Loop	Insight Loop
Stauffer	Counter Clockwise Thinking	Clockwise Thinking

2.0 DESIGN OF THE INSTRUMENT

2.1 Mindset

A key question in the design of this instrument was: what exactly does it measure? Is personal innovativeness a skill, an attribute, an attitude or perhaps a type of personal expertise? After considerable thought and investigation, I concluded that the answer is it is a *mindset*, mindset as described by Educational Psychologist Carol Dweck [8]. Dweck defines mindset as an "implicit theory" [19]. That is an assumption or belief that is not necessarily conscious, but that impacts how someone behaves, makes decisions and forms preferences. This approach is explicitly not an attempt to measure someone's fundamental personality or other innate traits such as IQ or multiple intelligences [20]. To the degree that we become conscious of our mindset, it is discretionary. It is something we can choose to change. So an innate traits model is not appropriate.[1]

Mindset is analogous to a computer operating system. It functions in the background, largely invisible to the user, and yet how it is designed has a profound effect on how the computer and all of its applications operate. So it is with a person's mindset—a personal operating system, made up of many individual files or *mental models*.

Another way to think about mindset is as a personal paradigm. Thomas Kuhn [21] adopted the term paradigm to describe the collection of theories and assumptions that scientists within a particular discipline share that guides their research, determines the questions they seek to answer and what problems they need to solve. Mindset serves this same function for each of us as individuals. It guides our actions and choices. But we are often not fully aware of the mindset we hold. We do not always realize what assumptions we are making.

Dweck's usage is similar to the concept of *theories-of-action* as articulated by Argyris and Schön [14] when they talk about the mental maps that guide someone's actions. They argue that we have frequently subconscious *theories-in-use* that actually guide our actions, and we also have what are sometimes very different *espoused theories* that we use to explain our actions to ourselves and others. Mindset as the term is used here refers to a person's *theories-in-use*.

Dweck [22] and her colleagues have spent decades researching the impact that mindset has on student performance, by comparing alternative theories of intelligence. One she calls a *fixed mindset* is the belief that our intelligence is innate and unchanging, something we have no ability to determine. The other she calls a *growth mindset*, the belief that with effort we can change our intelligence; we can make ourselves smarter. Dweck devised a survey that reveals which mindset a student prefers. She found that that preference can have a powerful effect on how a student performs, what risks the student is willing to take, what they value and believe about themselves and others, how they behave and compete and often whether or not they like school. She has found that mindset can be changed, and that significant changes in attitudes and behavior can occur as a result. In other words, our mindset is discretionary. It's something we have the power to revise if we want to—when we become aware of what our mindset is.

[1] In working with the instrument created here, the author has found that how a person scores is generally stable over time, but that with intervention (e.g. coaching or a workshop) someone's scores may shift dramatically.

Dweck's contribution is much more than her findings about a fixed vs. growth mindset; it is the concept of mindset itself as she has articulated it, and what she has demonstrated about its impact and its malleability. Her work has focused on two very specific implicit theories but it is easy to imagine how this same construct might apply in other contexts. I have applied her concept of mindset as implicit theories to innovativeness.

In the way they are used here, the terms *mindset, mental model, mental maps, implicit theory* and *paradigm* are largely interchangeable. However, I treat them as a hierarchy, one in which individual mental models (or implicit theories) combine to create an overall mindset (or personal paradigm). In this way, mindset describes following the Innovation or Status Quo Cycles and mental models describe the more specific phases or other components of those cycles. As Dweck and her co-researchers [19] have observed, "Because people's theories are largely implicit or poorly articulated, systematic effort is required on the part of behavioral scientists to identify them and to map out their effects." (p. 267) This is such an effort.

2.2 Human Innovativeness

The first step in designing this instrument was to translate the characteristics of the Innovation Cycle, which characterizes all forms of innovation—human and non-human, into a human cognitive/behavioral framework, to ask: what would it look like for someone to follow each of these two patterns? This is something I explored in, *Thinking Clockwise, A Field Guide for the Innovative Leader* [9].

Because the Innovation Cycle and the Status Quo Cycle mirror each other, they pose a series of tradeoffs much like the fixed vs. growth tradeoff that Dweck has investigated. This makes it possible to pose alternatives or choices that gauge whether someone has a tendency to follow the Innovation Cycle/Insight Loop, or the Status Quo Cycle/Knowledge Loop and the strength of that preference. So I began this design by imagining what those alternatives might be within each of the four phases that these two patterns share.

2.2.1 Idea Phase

The Idea Phase is about where we get our ideas and generate new possibilities. For the Idea Phase, following the Innovation Cycle would mean being comfortable using one's imagination, welcoming novelty and actively considering new possibilities. Alternatively, someone who favored the Status Quo Cycle in the Idea Phase would have a preference for what they already know and believe to be true. They would be skeptical of new possibilities and less interested in newness and novelty. So the Idea Phase poses a choice or trade-off between knowledge and imagination, to *know* or to *imagine*.

2.2.2 Action Phase

The Action Phase is about what we do with our ideas. In the Action Phase, favoring the Innovation Cycle means having the courage and desire to try new things, to explore and experiment. Favoring the Status Quo Cycle means preferring to apply existing knowledge and established processes and procedures. So the Action Phase poses a choice between application and experimentation, to *apply* or to *explore*.

2.2.3 Reality Phase

The Reality Phase is about what happens to us—the proverbial "reality check." So from a cognitive and behavioral perspective, it's about how we detect that, how we *observe*. Someone following the Innovation Cycle is intent on getting the most accurate data possible, welcoming negative as well as positive feedback, and actively looking for the exceptional and unexpected. Someone following the Status Quo Cycle is looking for validation, the assurance that the course of action they followed was a good one and has accomplished their objectives. Exceptions are treated as deviations that need to be corrected. The Reality Phase poses a choice between seeking validation, and challenging ourselves and the data to assure its accuracy and find the unexpected, to *validate* or to *challenge*.

2.2.4 Feedback Phase

The Feedback Phase is where we process the data we receive. It is often unclear why something does or does not work or exactly what the data is telling us, so someone following the Innovation Cycle is slow to draw conclusions, explores multiple interpretations and welcomes alternative perspectives.

They seek to reach new insights and make discoveries. Someone following the Status Quo Cycle is more likely to slot that information into their preexisting mental framework and strive to confirm what they already believe to be true. So the Feedback Phase poses a choice between seeking reinforcement and seeking to make discoveries, to *reinforce* or to *discover*.

Each phase can be represented as a spectrum of degrees of preference for these pairings or polarities that operationalize the Innovation and Status Quo Cycles. (Table 2) Posing choices in this way is important. Most of us would probably say we value aspects of both cycles, that we like knowledge *and* imagination, or both reinforcement and discovery. It's only by playing these alternatives off against each other that it is possible to gauge their *relative* importance to someone.

These alternatives and the items created to reflect them need not be opposites, nor mutually exclusive, but rather meaningful choices. Like choosing a flavor of ice cream, one does not have to dislike one to prefer another. For example in the Idea Phase, favoring imagination does not require that someone neglect knowledge or vice versa. What's important is the difference between the two, the degree to which someone favors one over the other.

Table 2. Choice Spectrums

Mindset	**Idea Phase**	**Action Phase**	**Reality Phase**	**Feedback Phase**
Trade-off Between	Generating Possibilities	Executing Possibilities	Experiencing Reality	Processing Feedback
Iterative Thinking	Imagine	Explore	Challenge	Discover
↕	↕	↕	↕	↕
Linear Thinking	Know	Apply	Validate	Reinforce

2.3 Creating Dimensions

As I began formulating possible items to represent the tradeoff between these two different mindsets, I realized they tended to fall into three categories.

1) Some items were about conscious personal *beliefs*, such as:
   When I'm taking action, it's important to… Use a proven approach. (or) **Explore possibilities.**
2) Some items described personal *values***:**
   I place greater value on… Definite answers. (or) **Multiple interpretations.**
3) Some items were self-reported *behaviors*:
   I find it easy to make thorough observations. (Agree…Disagree)

These three categories correspond in a general way to Dweck's findings that students' beliefs, values and behaviors reflect their underlying mindsets. Items were sorted into three profiles—a Cognitive (beliefs) Profile, a Values Profile and a Behavior Profile—for each of the four phases, for a total of 12 dimensions. (Table 3)

Table 3. Dimensions Matrix

	Idea Phase	**Action Phase**	**Reality Phase**	**Feedback Phase**
Cognitive Profile	Idea Cognitive	Action Cognitive	Reality Cognitive	Feedback Cognitive
Values Profile	Idea Values	Action Values	Reality Values	Feedback Values
Behavior Profile	Idea Behavior	Action Behavior	Reality Behavior	Feedback Behavior

← 12 Dimensions

This design strategy was intended to provide an indication of someone's overall mindset, and reveal details about a variety of personal beliefs, values and behaviors in the various phases, that reflect that mindset. This is like measuring someone's skill level at a variety of athletic activities—running, jumping, throwing—as a way to gauge their athleticism. Each measurement provides information about the overall capability, while also providing specific information about some particular tendency. That way, this instrument would serve as a developmental tool that provides detailed guidance to someone wanting to become more innovative, by revealing the specific adjustments they may need to make.

Innovativeness, as defined here, is not a single unidimensional attribute but rather a constellation of attributes that correspond to the Innovation Cycle. Developing multiple scales has two primary benefits; 1) it makes possible unidimensional measurements so these specific attributes can be cleanly measured and distinguished, and 2) it "samples" a person's preferences between these two patterns in multiple ways. Composites of these scales should then produce a more accurate indication of someone's overall tendency than any one scale may reveal.

This is not to say that the boundaries between these 12 dimensions are crisp bright lines. For example, imagination plays a role in all of the phases of the Innovation Cycle, not just the Idea Phase. In the Action Phase, our imagination helps us anticipate potential problems. In the Reality Phase, what one has not imagined as a possible outcome may be overlooked because it's not expected. In the Feedback Phase, being imaginative in one's interpretations is a productive strategy for gaining new insights. The Innovation Cycle is a dynamic multi-faceted process that these dimensions are designed to approximate.

3.0 MEASUREMENT METHODOLOGY

Rasch Measurement was used to calibrate this instrument. It was selected not only for its utility, but for its philosophy and methodology.

In this article and in the article explaining the Valuable Novelty Theory of innovation, I have drawn on the insights of Thomas Kuhn and here again his thinking is relevant. Kuhn [23] contends that the impression that even many scientists hold that empirical data is the ultimate arbiter of scientific validity is misleading. He observes that throughout the history of science, measurements have been as much a creation of theory as a test of it, that while measurement informs theory, theory informs measurement. Kuhn writes:

> "Many of the early experiments involving thermometers read like investigations *of* the new instrument rather than like investigations *with* it. How could anything else have been the case during a period when it was totally unclear what the thermometer measured?" (p. 188-189)

It takes only a moment's reflection to imagine how difficult it must have been to create a device such as a thermometer when there were no thermometers to compare it to, or to use for calibration. Creating a measure of a mindset of innovativeness poses a similar challenge. If the goal was to just provide criteria for sorting people into categories, as many existing instruments do, or compare people within a particular group (less than/more than), this problem might disappear. But the goal here was to create a universal metric that could be replicated and used to compare degrees of innovativeness between individuals, groups, organizations and perhaps even cultures…an innovativeness thermometer.

The Rasch methodology reflects Kuhn's conceptualization of the science of measurement, and Rasch measurement is probabilistic, just as the Innovation Cycle is probabilistic. It is worth noting that the dynamic interplay between theory and data that Kuhn observes is an example of the Innovation Cycle at work…as is the creation of this instrument.

Georg Rasch [24] proposed a model for calibrating the relative difficulty of dichotomous (right/wrong, true/false) items on a test, by determining the probability that a student of some particular ability would correctly answer an item of some particular difficulty. The greater the students level of skill, the higher the probability of getting an item correct; the higher the item's level of difficulty, the lower the probability of a student correctly answering it. A common analogy is a high jumper. The more capable the athlete, the greater the probability of clearing a bar of some particular height; the higher (more difficult) the bar, the lower the probability of someone of some particular skill level clearing it. When an item's difficulty (height of the bar) exactly matches a student's (athlete's) ability, the probability of a correct answer (clearing the bar) is .5 or 50%. As a student's ability increases, or the item's difficulty falls, the probability of success increases toward 1 or 100%. As a student's ability decreases or the item's difficulty rises, the probability of success decreases toward 0.

The model used here is one that David Andrich [25] subsequently developed based on Rasch's dichotomous model, to accommodate Likert-type items and other types of rating scales. This *polytomous Rasch model* makes it possible to create, test and calibrate measures (Likert type and semantic differential) that reflect degrees of preference between the Innovation Cycle and the Status Quo Cycle.

There are several crucial benefits that the Rasch methodology provides.

3.1 Linear Unit of Measurement

Rasch calculates probabilities by performing log normal calculations to produce log odds units or logits that serve the same function as degrees on a temperature scale. Logits measure the change in the probability of being innovative (in this application). One logit is the distance along the "scale" of the innovativeness variable that increases or decreases the odds of someone following (in this case) the Innovation Cycle in that phase and profile by a factor of 2.718.., the value of "e", the base of "natural" or Napierian logarithms [26]. In this way, logits provide a uniform additive unit of measurement.

3.2 Item Selection

The central challenge here was not to find a measurement that fits a predetermined set of indicators, but to identify the indicators that provide the best measurement. Rasch methodology provides a rigorous way to select such items. It enables a hypothetical deductive approach of imagining possible items and then determining whether those items are a good fit. This is a fundamental difference between Rasch and item response theory (IRT) [27]. Rasch is sometimes described as one implementation of IRT and they are mathematically equivalent, but standard practice with IRT is to select the specific model that best fits the data at hand. The data is assumed to be valid for measurement purposes. In the creation of a new instrument such as this, that was not a well-founded assumption. Some items were no doubt more indicative than others. The impact of a number of confounding factors including how items will be interpreted, social acceptability and potential multi-dimensionality were difficult to discern. Unlike an academic examination, there were no specific skills that must be tested. So the individual items were less important than whether those items were good indicators of the mental models and mindset they were intended to measure. To return to the thermometer analogy, it didn't make sense to select raw materials (items) and then figure out what kind of thermometer (measurement scale) to make. What made sense was determining the specific kind of thermometer needed and then identifying the appropriate materials to construct it.

3.3 Relative Item Strength

It's unlikely that any two items are equal in their ability to reflect the hypothesized variable. (Just as no two words are equal reflections of someone's ability to spell.) Rasch measurement accounts for this by making probabilistic calculations of the relative strength or *fit* of each item as an indicator.

3.4 Directionality

Rasch analysis looks for a clear directional signal among the items, from least difficult/preferred to most difficult/preferred [28]. This is essential because an instrument like this must be able to distinguish between a tendency toward the Innovation Cycle and a tendency toward the Status Quo Cycle. This tradeoff can be represented on the same scale because more of one preference is theorized to be less of the other and vice-versa—but the direction of the scale must be clear and consistent. Rasch methodology determines whether the measures align in such a sequence—a test of *construct validity* [1].

3.5 Parameter Separation

A reliable instrument should accurately measure a person's innovativeness at the same level no matter how the items are distributed and no matter who is taking the assessment. In other words, a high jumper should have the same probability of clearing a specific height no matter how many ways the bar is placed and the placement of the bar should pose the same degree of difficulty no matter who is doing the jumping. This parameter separation is required in order for proposed items to fit the Rasch model [29], something that is not required with Item Response Theory and True Score Theory [30].

3.6 Hypothesis Testing

Rasch methodology effectively tests the underlying theory and the Phase and Profile constructs by determining whether or not items can be created that provide unidimensional unidirectional measures that are consistent with the theory. If items could not be found that fit the Rasch model, and its component dimensions, the soundness of the theory would be in doubt.

3.7 Parallels with Theory

Rasch methodology parallels the Valuable Novelty Theory in another intriguing way. The theory posits that the Innovation Cycle and the Status Quo Cycle are each symmetrical. That is to say that the two sides of each of these feedback loops are in a sense interchangeable. An entity striving to innovate must somehow accommodate its environment, but it also impacts that environment and therefor serves as part of the reality that other entities must accommodate. A living organism is part of and partially defines the larger ecosystem. A successful new product impacts other existing and potential products. One company's growth impacts other companies' ability to grow. So the same entity can logically be placed on the left side of the cycle as a source of new possibilities, or on the right as a source of feedback. Both the Status Quo and Innovation Cycles describe this kind of mutual influence.

Rasch methodology does something similar with persons and items [33]. While there are obvious practical differences between those whose attributes are being measured and the items used to measure them, they are mathematically interchangeable in Rasch. To return to the analogy of high jumping, there are two ways to impact the probability of a successful jump; 1) change the capabilities of the jumper, and 2) move the bar. Either of these approaches work and the relationship between the two is what Rasch models describe. Which is which makes no difference mathematically. The Winsteps program [31], which performs Rasch analysis and was used here, works equally well with persons and items transposed.

Another shared feature of the Rasch methodology and the Innovation Cycle is that both are probabilistic models. These parallels between theory and methodology suggest that Rasch measurement is not only a good choice for the design of this instrument, but may point to a way to mathematically represent the Innovation Cycle itself. That's beyond the scope of this article but it deserves further exploration.

4.0 CREATING AND TESTING THE INSTRUMENT

4.1 Pilot Study

4.1.1 Survey Items

A total of 159 proposed items were created and sorted among the 12 dimensions. All of them were one of two types:

Likert Type Scale – Nine Point
These were a mix of iterative and linear indicators. They used either a frequency scale (Never, Rarely, Sometimes, Often, Always) or an agreement scale (Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree)

Semantic Differential – Seven Point
These were anchored on one end by an Iterative Thinking attribute and on the other end by a Linear Thinking attribute.

4.1.2 Participants

Five diverse organizations were recruited to participate in the pilot study: a research and development group within Medtronic Cardiovascular, an IT unit within Thomson Reuters, a treatment and rehabilitation unit of Alina Health Systems, Synovis Life Technologies and Spring Lake Park-Blaine-Moundsview Fire Department. Participants ranged from front-line employees and their supervisors to high-level managers and senior executives. A total of 366 people were enrolled by their respective organizations and allowed to opt out. There were 109 people who chose not to participate, leaving an *n* of 257, a participation rate of 70%. Of those who participated, 141 were men (55%) and 116 were women (45%). All participants completed the instrument online.

4.1.3 Data Analysis Software

WINSTEPS Version 3.81.0 [28] was the Rasch measurement software used to perform the data analysis.

4.1.4 Grouping

Where there was a mix of Likert-type and semantic differential rating scales within a dimension, they were first sorted between these types. Then within each type, the item rating scale categories were collapsed into groups based on the response patterns, using the Rasch fit statistics and category structure as a guide. This was done to satisfy the requirement of the polytomous Rasch model that rating scale categories increase monotonically from lowest to highest preference [32]. The transitions (Andrich thresholds) from one degree of preference to the next need to be sequential never reversed in their relative probabilities. This is an important step in the creation of ordered response measures with

Rasch, one that is analogous to assuring that the glass tube of a thermometer has a uniform diameter, so that the mercury rises in a way that's continuous, consistent and additive.

4.1.5 Item Fit

With Rasch measurement, there are two primary ways that fit to the model is measured: *Infit* and *Outfit* [33,34]. Outfit is based on the conventional chi-square statistic. For each item, the residual for each person that responded to an item is squared, then the values for all persons are summed and divided by the number of persons. The same calculation is made for each person based on all the items they encounter. Outfit can be sensitive to outliers or unusual data values, so Infit is also calculated. Infit is a chi-square statistic weighted according to its model variance, making it more representative of localized patterns in the data.

Optimal Infit or Outfit to the Rasch model is indicated by a mean square fit statistic of 1.0 Mean-square values below 1.0 indicate over fit (more predictable than the model expects) while values over 1.0 indicate under fit (less predictable than the model expects). Items that under fit the model (high mean-square values) introduce noise that can lower reliability, while items that over fit the model (low mean-square values) may produce inflated reliability values.

Exact fit of the data to the Rasch model (or any other model) is not a realistic expectation. To be considered useful for measurement purposes, there are no hard fast rules on mean-square values, but Wright and Linacre [36] recommend that mean-square item fit statistics fall between 0.6 and 1.4 for rating scale instruments of this type, and within a range of 0.8 to 1.2 for high stakes instruments. The 0.6-1.4 criteria was used here, with Infit values (chi-square statistics) for all items, and Outfit values for all but 2 items, across all 12 scales falling within that range. Seven of the 12 scales also met the more stringent high stakes criteria.

4.1.6 Reliabilities

As expected, the Rasch analysis found that some items did not fit the model and these indicators were dropped, leaving 77 items ranging from 3 to 10 items per dimension. The reliabilities for the remaining items and persons are shown in Table 4. Per WINSTEPS guidelines, model reliabilities are reported. [31] *Person reliability* measures how consistently the same persons would score when given another set of items that measure the same construct. This is comparable to the *test reliability* estimated in Classical Test Theory with Cronbach Alpha and KR-20. *Item reliability* is a metric used in Rasch analysis to measure how consistently the same items would perform when given to another similar group of persons. (Ibid)

With Rasch analysis, person reliability should be approximately 0.8 or higher, and item reliability should be above 0.9 [35]. Person reliabilities were sufficient to categorize people into two levels (more/less innovative or linear/iterative by phase) as this instrument is designed to do. The one exception was the Feedback Behavior dimension with a person reliability 0.71 and separation of 1.55. All item reliabilities were more than sufficient to show a clear item hierarchy and therefore a degree of *construct validity* for the measures in each dimension [1].

Table 4. Rasch Analysis Results – Pilot Study

Dimension		Model Reliability / Separation		Items	
Phase	Profile	Person	Item	Tried	Kept
Idea	Cognitive	.86 / 2.47	.98 / 6.98	11	5
	Values	.83 / 2.21	.99 / 9.20	7	6
	Behavior	.85 / 2.42	.98 / 6.43	19	6
Action	Cognitive	.88 / 2.72	.98 / 6.54	10	7
	Values	.87 / 2.60	.95 / 4.25	9	6
	Behavior	.82 / 2.13	.99 / 12.18	25	10
Reality	Cognitive	.83 / 2.19	.97 / 9.54	8	5
	Values	.81 / 2.04	.97 / 5.47	5	3
	Behavior	.80 / 1.97	.99 / 9.15	22	6
Feedback	Cognitive	.80 / 2.01	.99 / 9.31	8	6
	Values	.82 / 2.12	.96 / 4.82	10	7
	Behavior	.71 / 1.55	.97 / 5.33	25	10
Total				159	77

4.1.7 Scaling Dimensions and Their Composites

Determining how to best combine these dimensions into the various composite Phases and Profiles and into an overall score posed several challenges: 1) Correlations between dimensions were low to moderate (See post hoc analysis.), so they could not be analyzed together as composite scales without creating significant misfit and parameter bias [36]. 2) To be comparable, these dimensions would normally need to be standardized so that the logit measures are equivalent (much like Fahrenheit and Celsius temperatures). But that created some dimensions with ranges that were as much as twice the size of others. Mathematically, this gave some dimensions much more weight than others, something that is not theoretically justified. 3) Processing twelve dimensions exceeds the capabilities of most multidimensional Rasch models and it is not clear that it would eliminate misfit and avoid introducing bias [37]. A possible exception is a Monte Carlo method [38], which may be an option for future research.

The approach that was adopted in this pilot study and in subsequent analyses is based on the nature of what was being measured. These dimensions did not evaluate a person's absolute preference for constructs such as Imagination and Knowledge (which might both be high) but rather how one struck a *tradeoff between* the two. This tradeoff can be expressed as a ratio that can be compared across dimensions, regardless of the increments or range, as long as those ratios are based on measurements that are uniform along each scale—as the Rasch analysis provided. This approach did not accurately compare logit measures between dimensions but that is of little consequence because such comparisons have almost no intelligible meaning. It makes little sense to say that someone has a logit, or other unit of measurement, more of Knowledge than they have of Imagination. It makes even less sense to compare differing dimensions such as Knowledge Value and Feedback Behavior in this way. It does make sense to say someone prefers one construct over the alternative by some ratio and that preference is stronger for some pairings than for others.

So the measures for each dimension were rescaled into 100 units, [39] with 0 being the lowest possible score (favoring the Status Quo Cycle) and 100 being the highest (favoring the Innovation Cycle. In this way, each person's *relative* preference was expressed as a percentage of each dimension's operational range. Dimension scores were then combined into composite Phases and Profiles (Table 3) and an overall *Innovativeness Index* by calculating the mean of these percentages for the dimensions within each composite. This approach created meaningful composites while preserving the integrity and fit of each dimension. Implicit in this approach is that the preferences reflected in these dimensions, when combined, represent someone's overall preference between the Innovation and Status Quo cycles. The higher the Innovativeness Index, the more someone favors the Innovation Cycle; the lower the Index, the more someone favors the Status Quo Cycle.

This version of this instrument, now known as Innovator Mindset, has been in use for several years with more than a thousand evaluations taken by widely diverse participants. This has been done for research as well as in workshop and other educational settings in which attendees are interested in enhancing their personal innovativeness. Data has been gathered on experimental items for the various dimensions for the purpose of increasing the instrument's validity and reliability. This was done particularly for the Reality Cognitive dimension that had the fewest items that the Rasch analysis showed to be good indicators, and the Feedback Behavior dimension that showed the lowest person reliability.

4.2 First Iteration

4.2.1 Dataset, Reliability and Fit

With assistance from the Kauffman Foundation for Entrepreneurship, the Innovator Mindset instrument was given to 328 entrepreneurs [3] including 234 men and 94 women from 43 states and the District of Columbia. Business types, size, business models and sectors were diverse. Data from this group was used to do a second Rasch analysis on this instrument, including testing and calibrating all existing and experimental items. Reliabilities and separation values for the retained items and persons for each dimension are shown in Table 5.

Items were again required to have Infit and Outfit mean square values within the 0.6-1.4 range recommended by Wright and Linacre. [40]. To reduce noise in the calculations, persons with Infit and Outfit values that were above 2.0 (MNSQ) *and* statistically significant at $\alpha < .05$ (ZSTD statistic above 2.0) were dropped.

Overall, item reliabilities improved from the pilot study, while person reliabilities were a little lower. The weakest dimension reliability and biggest drop in reliability was Reality Behavior at 0.70 as compared to 0.80 for this dimension in the pilot study. This suggests that the difference is due more to the nature of the distribution of persons than to the strength of the items. As a group, entrepreneurs had higher average scores and a narrower range of scores on some dimensions than the pilot study group.

Two existing items were dropped from the analysis due to lack of fit to the model (but will be kept in the instrument survey for further evaluation). Two items that fell just below the minimum mean square values (Infit 0.58, Outfit 0.59) were also kept in pending further research. Of the additional items tested, 9 fit the model sufficiently well to be added, for a net total of 84 items.

Table 5. Rasch Analysis Results – First Iteration

Dimension		Model Reliability / Separation		Item Count	PCA Unexplained Variance in 1st Contrast Plot		DIF Female/Male
Phase	Profile	Person	Item		Eigenvalues	Percentage	Net Contrast
Idea	Cognitive	.78 / 1.86	.99 / 9.87	5	1.7	16.7 %	-0.04
	Values	.84 / 2.27	.99 / 8.39	6	1.6	11.6 %	0.10
	Behavior	.80 / 1.97	.99 / 10.37	5	1.8	14.3 %	0.07
Action	Cognitive	.83 / 2.18	.99 / 8.25	8	1.8	11.5 %	0.09
	Values	.84 / 2.28	.98 / 6.62	6	1.6	12.5 %	-0.02
	Behavior	.88 / 2.72	1.00 / 16.09	8	1.5	8.0 %	0.04
Reality	Cognitive	.83 / 2.22	.99 / 9.79	5	1.8	13.9 %	-0.41
	Values	.83 / 2.23	.97 / 5.71	6	1.6	12.7 %	0.04
	Behavior	.70 / 1.54	.99 / 8.59	11	1.7	10.0 %	0.50
Feedback	Cognitive	.84 / 2.29	.99 / 10.79	6	1.8	12.8 %	-0.09
	Values	.88 / 2.67	.96 / 5.07	9	1.6	9.1 %	0.71
	Behavior	.81 / 2.04	.99 / 8.65	9	1.6	9.8 %	-0.12
Total				84			

4.2.2 Dimensionality

Given the multifaceted nature of the Innovation and Status Quo Cycles and the human attributes they imply, a degree of multidimensionality was expected.

WINSTEPS does a principal components factor analysis (PCA) that unlike traditional factor analysis, does not seek to maximize commonalities among factors, but rather looks for variability or contrasts between opposing factors [41]. It does this by looking at both positive and negative loadings, and at residuals rather than observations. This is done to try "…to falsify the hypotheses that the residuals are random noise by finding the component that explains the largest amount of variance in the residuals." (Ibid)

PCA results showed that Eigenvalues for the unexplained variance in the first standardized residual contrast plot were within recommended guidelines (< 0.2) for all dimensions. So there do not appear to be any secondary dimensions large enough to undermine the validity of these measures.

4.2.3 Grouping and Scaling

As in the pilot study, items were sorted by type, and rating scale categories were again grouped based on the response patterns and model fit. Composite Phases, Profiles and an overall Innovativeness Index were again calculated.

4.2.4 Differential Item Functioning (DIF)

A gender DIF analysis was done using the Mantel test and the Rasch-Welch t-test. Only one item showed a statistically significant contrast ($\alpha = .01$). This item on the Reality Cognitive dimension favored males by 2.91 logits. However, when all DIF contrasts for that dimension were combined, males were favored by just .41 logits. DIF contrasts netted out to less than one logit on every dimension, so there does not appear to be any systematic gender bias. (Table 5)

4.2.5 Post Hoc Analysis

A post hoc analysis of these dimensions showed that they all correlated positively to each other. (Table 6)

- Pearson r correlations for comparisons between individual dimensions were low to moderate, with a mean of 0.51
- Dimensions compared within each phase (Idea, Action, Reality, Feedback) were somewhat higher with a mean value of 0.65
- Dimensions compared within each Profile (Cognitive, Values, Behavior) also showed a somewhat higher mean value of 0.68
- Dimensions correlated even more highly with the overall Innovativeness Index with a mean correlation of 0.74
- The composite Phases and Profiles, when compared to each other, also had significantly higher correlations with mean values of 0.71 and 0.76 respectively
- The highest correlations were found when comparing the composite Phase and Profile scores to the Innovative Index, with mean r values ranging from 0.88 and 0.92

Table 6. Correlations

Correlations	**Mean**	**Min.**	**Max.**
Dimension to Dimension	0.51	0.23	0.76
Dimensions Within Phases	0.65	0.33	0.90
—Idea Phase	0.63	0.33	0.86
—Action Phase	0.68	0.47	0.90
—Reality Phase	0.64	0.39	0.84
—Feedback Phase	0.67	0.45	0.88
Dimensions Within Profiles	0.68	0.48	0.89
—Cognitive Profile	0.66	0.53	0.81
—Values Profile	0.68	0.48	0.89
—Behavior Profile	0.70	0.53	0.82
Dimensions to Index	0.74	0.62	0.87
Phase to Phase	0.71	0.58	0.81
Profile to Profile	0.76	0.70	0.83
Phases to Index	0.88	0.86	0.91
Profiles to Index	0.92	0.88	0.94

These correlations are what one would expect if the Valuable Novelty theory is sound, and this instrument is performing as it should. They therefore provide further evidence of construct validity, showing that:

1. All dimensions point in the same direction (no negative correlations).
2. Each dimension is contributing unique information. (No two are redundant.)
3. Dimensions are more strongly correlated to the composites on average than they are to each other, and are most strongly correlated to the Innovativeness Index.
4. Correlations are stronger on average between Phases, Profiles and the Innovativeness Index than they are between composites and dimensions.
5. There appears to be a latent variable common to all dimensions.

4.2.6 Negative Correlations

An interesting aspect of this data was that a substantial number of persons showed strongly negative point measure correlations. For some dimensions, negative correlations were found for as many as a third of the persons and as strong as -.98. This was in contrast to item measures that formed a clear hierarchy that set the direction for the scales derived from them. Such negative person correlations suggest a possible violation of a key requirement of Rasch measurement: that there be a clear linear progression from least preferred to most preferred items. In another context this "reversal" of preferences would be counterintuitive, perhaps indicating that a math student found algebra easier than addition, or that a high jumper's success increased with the height of the bar rather than inverse to it. However, in this application it makes sense…and actually provides support for the Valuable Novelty Theory.

Since more Linear Thinking implies less Iterative Thinking and vice versa, someone whose mindset favors the Status Quo Cycle would not only score differently from someone whose mindset favors the Innovation Cycle, but would be expected to rank those preferences in reverse order. The fact that this occurs to such a substantial degree suggests that it is due to more than just individual idiosyncrasies. This would not invalidate the instrument if the item measures have a clear ranking (as they do). Removing all of the persons with negative correlations from each dimension tested this interpretation. That did no more than marginally increase person or item fit and for some dimensions it had no impact on fit to the model.

5.0 FINDINGS

The successful creation of this instrument and its fit to the polytomous Rasch model provides empirical support for the Valuable Novelty theory of innovation and innovativeness by demonstrating that:

1. *Clockwise or Iterative Thinking (following the Innovation Cycle) and Counter Clockwise or Linear Thinking (following the Status Quo Cycle) are prevalent mindsets that can be detected and measured.*
2. *These two mindsets are not mutually exclusive, but it is possible to distinguish between them.*
3. *Each of these mindsets prompts a variety of beliefs, values and behaviors that are logically connected to one mindset or the other, such that these manifestations can be used to detect and measure the degree to which a person holds one mindset or the other.*

The pilot study post hoc analysis that showed positive correlations among the dimensions, supports the theoretical assertion that:

1. *The Status Quo and Innovation Cycles are self-reinforcing ; more of one dimension implies more of other dimensions for both patterns.*

The finding that such a substantial number of persons rank some of their preferences in “reverse” order supports another aspect of the Valuable Novelty Theory, that:

2. *The Status Quo and Innovation Cycles logically flow in opposite directions.*

6.0 DISCUSSION

This instrument’s design expanded on the Valuable Novelty Theory by defining it in terms of human-fostered innovation (or the lack of it). Further research is needed to continue to refine this instrument, including the creation of more items and more varied items to enhance person (test) reliability. Further investigation may include exploring other potential analytical models such as an unfolding Rasch model, a mixed Rasch model and one or more multidimensional Rasch models. Research is also needed to determine whether these alternative mindsets can be linked to real world value creation—the purpose of the third article in this series [3].

To achieve a truly universal metric for innovativeness, a single standardized scale needs to be developed, one that equates these dimensions across multiple versions of the instrument. Comparing each dimension to itself should avoid the problems encountered in trying to equate the dimensions to each other, and should make the composite Phases, Profiles and the Innovativeness Index comparable as well. Ideally, this should be done with a diverse and representative national or perhaps international dataset. This is a logical next step now that the instrument has shown strong internal (construct) and external validity. (See companion paper.) [3].

In addition to an overall metric, the various Phases and Profiles provide a richly detailed picture of someone’s mindset and what adjustments can be made to enhance innovativeness. Individuals may show considerable dissonance among these dimensions. Someone may not choose between Imagination and Knowledge to the same degree, or even in the same theoretical direction, that they choose between Challenge (seeking the best data) and Validation (confirming that they’re right). Even within each Phase, a person’s values, beliefs and behaviors were not necessarily aligned. This instrument should be a useful tool for revealing someone’s mindset and its impact on their capacity to innovate. It lays a foundation for greater self-awareness and personal development.

This instrument also makes it possible to identify individuals with a high likelihood of being innovative, a capability that should be of value to organizations in their talent recruitment and to early stage investors as part of due diligence. It has potential applications in a range of educational settings as a way to gauge whether students are becoming more or less innovative as a result of some curriculum or intervention. And, it can serve as a research tool to measure and compare innovativeness in many diverse settings, disciplines, professions, organizations, demographic groups and cultures.

ACKNOWLEDGEMENTS

Thank you to Paul Bergmann for introducing me to Rasch methodology and doing the analysis for the pilot study, and to Everett Smith for his guidance and review of my application of Rasch. Thank you to Bettina von Stamm for her encouragement and editorial review.

REFERENCES

[1] Wright Benjamin D., Masters, Geofferey N. *Rating Scale Analysis* [Book]. - Chicago : Institute for Objective Measurement, Inc., 1982.

[2] Stauffer Dennis A Valuable Novelty: A Proposed General Theory of Innovation and Innovativeness [Journal]. - [s.l.] : *IJIS*, 2015. - 3 : Vol. 7.

[3] Stauffer Dennis A. Innovativeness as a Predictor of Entrepreneurial Value Creation, accepted, [Journal]. - [s.l.] : *IJIS*, 2016. - 1 : Vol. 8.

[4] Plucker Jonathan A and Renzulli Joseph S Psychometric Approaches to the Study of Human Creativity [Book Section] // *Handbook of Creativity* / ed. Sternbreg Robert J. - Cambridge : Cambridge University Press, 1990.

[5] Myers Isabel Briggs *The Myers-Briggs Type Indicator Manual* [Book]. - Palo Alto, CA : Consulting Psychologists Press, 1962.

[6] Kirton Michael Adaptors and Innovators: A Discription and Measure [Journal] // *Journal of Applied Psychology*. - 1976. - 6 : Vol. 61. - pp. 622-629.

[7] Puccio Girard J., *FourSight: The Breakthrough Thinking Profile – Presenter's Guide and Technical Manual*, [Book], - Evanston : THinc Communications, 2002

[8] Dweck Carol S. Mindset, *The New Psychology of Success* [Book]. - New York : Ballantine Books, 2006.

[9] Stauffer Dennis A. *Thinking Clockwise, A Field Guide for the Innovative Leader* [Book]. - Minneapolis : MinneApplePress, 2005.

[10] Whitehead Alfred N. *Modes of Thought* [Book]. - New York : Free Press, 1938.

[11] Wertheimer Max *Productive Thinking* [Book]. - New York : Harper and Row, 1945.

[12] Piaget Jean *The Origins of Intelligence in Children* [Book]. - New York : International University Press, 1952.

[13] Cattell Raymond B. Theory of Fluid and Crystalized Intelligence: A Crucial Experiment [Journal] // *Journal of Educational Psychiology*. - 1963. - Vol. 54. - pp. 1-22.

[14] Argyris M Chris and Schön Donald *Theory in Practice. Increasing professional effectiveness* [Book]. - San Francisco : Josey-Bass, 1974.

[15] Bateson Gregory *Naven* [Book]. - Stanford, California: Stanford University Press, 1958. - 2nd.

[16] Ashby W Ross *Design for a Brain* [Book]. - New York : John Wiley and Sons, 1960. - 2nd.

[17] Kahneman Daniel *Thinking Fast and Slow* [Book]. - New York : Farrar, Strauss and Giroux, 2011.

[18] Evans Jonathan St. B. T. and Stanovich Keith E *Dual-Process Theories of Higher Cognition: Advancing the Debate* [Journal]. - Washington, D.C. : Sage, 2013. - 3 : Vol. 38. - pp. 223-241.

[19] Dweck Carol S, Chui Chi-yue and Hong Ying-yi Implicit Theories and Their Role in judgements and Reactions: A World from Two Perspectives [Journal] // *Psychological Inquir*y. - 1995. - 4 : Vol. 6. - pp. 267-285.

[20] Gardner Howard *Frames of Mind, The Theory of Multiple Intelligences* [Book]. - New York : Basic Books, 2011.

[21] Kuhn Thomas S. *The Structure of Scientific Revolutions, Second Edition, Enlarged* [Book]. - Chicago : University of Chicago Press, 1962, 1971.

[22] Dweck Carol S *Self-Theories: Their Role in Motivation, Personality and Development* [Book]. - Philadelphia : Taylor and Francis, 2000.

[23] Kuhn Thomas S *The Function of Measurement in Modern Physical Science* [Journal] // Isis. - [s.l.] : The University of Chicago Press, 1961. - 2 : Vol. 52. - pp. 161-193.

[24] Rasch Georg *Probabalistic models for some intelligence and achievemet tests* [Book]. - Copenhagen : Danish Institute for Educational Research, 1960. - (Expanded edition, 1980 Chicago: Univeristy of Chicago Press.

[25] Andrich David *A rating formulation for ordered response categories* [Journal] // Psychometrika. - 1978. - Vol. 43. - pp. 561-574.

[26] Linacre John M and Wright Benjamin D The "Length" of a Logit [Journal] // *Rasch Measurement Transactions*. - 1989. - 2 : Vol. 3. - pp. 54-55.

[27] Andrich David Controversy and the Rasch Model: A Characteristic of Incompatible Paradigms [Book Section] // *Introduction to Rasch Measurement* / ed. Smith Everett V., Jr and Smith Richard M. - Maple Grove : JAM Press, 2004.

[28] Stone Mark H Substantive Scale Construction [Book Section] // I*ntroduction to Rasch Measurement* / ed. Smith Everett V., Jr and Smith Richard M. - Maple Grove : JAM Press, 2004.

[29] Rasch Georg On General Laws and the Meaning of Measurement in Psychology [Conference] // *Proc. Fourth Berkeley Symp. on Math. Statist. and Prob.*. - Berkeley : University of California Press, 1961. - Vol. 4. - pp. 321-333.

[30] Smith Everett V., Jr Evidence for the Reliability of Measures and Validity of Measure Interpretation: A Rasch Measurement Perspective [Book Section] // *Introduction to Rasch Measurement* / ed. Smith Everett V., Jr and Smith Richard M. - Maple Grove : JAM Press, 2004.

[31] Linacre John M WINSTEPS 3.81.0. - 2014.

[32] Andrich David A Thresholds, Steps and Rating Scale Conceptualization [Journal] // *Rasch Measurement Transactions*. - 1998. - 3 : Vol. 12. - pp. 648-649.

[33] Bond Trevor G and Fox Christine M *Applying the Rasch Model, Fundamental Measurement in the Human Sciences* [Book]. - New York : Routledge, 2007. - 2nd.

[34] Linacre John M *Misfit diagnosis: infit outfit mean-square standardized* // WINSTEPS 3.81.0. - 2014b.

[35] Wright Benjamin D Reliability and Separation [Journal] // *Rasch Measurement Transactions*. - 1996. - 4 : Vol. 9. - p. 472.

[36] Folk Valerie Greaud and Green Bert F Adaptive Estimation When the Unidimensionality Assumption of IRT Is Violated. [Journal] // *Applied Psychological Measurement*. - [s.l.] : ERIC (U.S. Dept. of Education), 1989. - 4 : Vol. 13. - pp. 373-89.

[37] Adams Raymond J, Wilson Mark and Wang Wen-chung The Multidimensional Random Coefficients Multinomial Logit Model [Journal] // *Applied Psychological Measurement*. - [s.l.] : Sage Publications, Inc., 1997. - 1 : Vol. 21. - pp. 1-23.

[38] Volodin Nikolai and Adams Raymond J The Estimation of Polytomous Item Response Models with Many Dimensions [Online] // ACEReSearch. - Australian Council for Educational Research, 12 2002. - 12 2015. - http://research.acer.edu.au/ar_misc/14.

[39] Linacre John M *User-friendly rescaling: zero point and unit* // WINSTEPS 3.81.0. - 2014c.

[40] Wright Benjamin D and Linacre John M Reasonable mean-square fit values [Journal] // *Rasch Measurement Transactions*. - 1994. - 3 : Vol. 8. - p. 370.

[41] Linacre John M *Dimensionality: Contrasts and Variances* // WINSTEPS 3.81.0 Documentation. - 2014a.