Innovator Mindset Technical Report

An explanation of the scientific rigor behind Innovator Mindset

This technical report explains the statistical methodologies used to create and validate the Innovator Mindset assessment instrument. A decision to use any assessment instrument should be based upon careful consideration of the research behind it, including its reliability and validity when used appropriately in the intended context. Not all assessments meet the same scientific standards.

Item Response Theory vs. Factor Analysis – Why it Matters

The Innovator Mindset assessment instrument (IM) was created using Rasch analysis, a type of Item Response Theory (IRT). While IRT (also called *latent trait analysis*) is widely viewed as a superior approach, many popular assessments do not have this same level of rigor.

Item Response Theory is a scientific paradigm based on the observation that each of the items on a survey will be of somewhat different difficulty, power or predictive capability. This contrasts with Classical Test Theory (CTT) which assumes that all items are of equal value. Traditional Factor Analysis (FA) makes this same assumption of equivalence. In reality, this is almost certainly not the case, whether the instrument is a second grade spelling test or a sophisticated measure of personality. Critics of CTT and FA have argued that this approach is like trying to compare the athletic performances of two high jumpers, by counting how many times they jump over a bar, but without measuring the height of those bars.

Both IRT and CTT/FA are used to determine three things about a measurement instrument. Is it *valid, reliable* and *appropriate*? The validity of the instrument is determined by how well it measures what it claims to measure. Its reliability is determined by how dependable it is at consistently doing that. Its appropriateness is determined by how well it serves its intended purpose. CTT/FA evaluate these widely recognized standards at the level of the instrument. IRT evaluates them at the level of the instrument and also analyzes how well *each item* works, or *fits* the design and purpose of the instrument.

IRT originated as a way to more accurately calibrate educational testing, and to reliably compare the scores of students on one test, to the scores of other students on a different test. It calculates the probability of a correct (or predictive) response to an item by considering the relationship between the difficulty of the item and the capability of the person responding to that item. So it considers how many items a student answers correctly *and* adjusts the score based on the difficulty of those items.

The added rigor that IRT provides has led to its adoption for many instruments used for high stakes decisions, especially in academia. IRT makes it possible to create item banks, so different items may appear on different versions of the same instrument. Yet scores on those different versions can still be fairly compared. IRT is necessary to enable computerized adaptive testing (CAT). CAT adjusts the difficulty of the items provided, based on the examinee's ability as reflected in prior responses. IRT analysis is used in medical clinical research because of its high level of precision.

Another advantage of IRT is that it does a better job of accounting for biases like social acceptability. When bias is present in the population that an instrument is designed to assess, IRT tends to adjust for that bias. When an item is overwhelmed by bias, it no longer "fits" and under IRT analysis it would be dropped from the assessment. This means that an IRT-designed instrument is less likely to be undermined by bias. Think of someone being asked if they are a Neo-Nazi. This is so socially unacceptable in most circles that few people will admit to it even when it may be true. So when someone does admit to it, we know they must be really committed. The lack of social acceptability becomes one way to gauge the strength of those convictions.

CTT is easier to administer and explain than IRT. Computer software is needed to implement IRT. However this is not usually a concern because most assessment instruments are now computer-based anyway (providing consistency of administration that enhances validity and reliability). The following table categorizes many widely used assessment instruments according to whether they have been designed using IRT. Failing to use IRT does not mean the instrument is not valid, reliable and appropriate. But it does mean that those attributes have not been evaluated with the high level of rigor and precision that IRT requires. So they may be less accurate and may not be as appropriate for making high stakes decisions (e.g. admission, hiring or advancement).

Analysis used on some widely available assessment instruments	
Item Response Theory	Non-IRT Factor or other analysis
SAT - Scholastic Aptitude Test PSAT - Preliminary Scholastic Aptitude Test ACT - American College Testing GRE - Graduate Records Examination GMAT - Graduate Management Admission Test LSAT - Law School Admissions Test IQ - Stanford-Binet Intelligence Measurement Innovator Mindset	Myers Briggs Type Indicator Insights Discovery Innovator's DNA Swarm Innovation Profiler Clifton StrengthsFinder Gallop Entrepreneurial Profile Entrepreneurial Mindset Profile Predictive Index Creatrix FourSight Kirton Adaptor Innovator ViaEdge Talentx7 DISC

Another important distinction between these two lists is that the assessments on the right side focus on things like personality traits, cognitive style and behaviors. Any *gaming* of such an assessment, by the examinee misrepresenting themselves, undermines its validity and reliability. The assessments in the left column measure things like knowledge, mindset and cognitive abilities. These are things that can be learned and developed. In order for someone to game one of these assessments, they would need to know the correct answers. That is what these assessments are trying to determine anyway. So gaming is less of a concern.

For example, when someone goes to law school, one of the outcomes is knowing the correct answers on the bar exam, but we don't consider that a problem. We assume that knowing those things increases the likelihood that they will be applied. Similarly, for someone to successfully game Innovator Mindset, the examinee needs to understand the underlying model. Teaching that model is one purpose of the Innovator Mindset approach.

Rasch Analysis

Innovator Mindset was created and validated using Rasch analysis, a specific type of IRT. Rasch is now widely used by many different researchers in a wide range of fields. The main difference between Rasch and other forms of IRT is in the philosophy that underpins it. Like

other statistical modeling approaches, IRT emphasizes the primacy of the data and making the model fit the data. Rasch turns this around and requires that the data fit the Rasch measurement criteria. Items that fail to meet this standard are rejected.

It is a common practice among psychologists to retain and score items that are only weakly supported by IRT or FA, when the assessment designer believes that those items still produce useful information. Innovator Mindset has not done this. All scored items in Innovator Mindset have been shown mathematically to fit the Rasch criteria.

Innovator Mindset is composed of 12 subscales or dimensions. Each of these twelve scales was analyzed separately using the Rasch model. These dimensions were then combined into *Phases*, *Profiles* and an overall *Innovativeness Index*. While these dimensions correlate to each other to varying degrees, they are each designed to measure a unique attribute. So the Innovativeness Index is not a single scale, but rather a measure of how much someone follows a larger pattern made up of those dimensions—the *Innovation Cycle*. This is analogous to measuring more than one skill or type of intelligence, or the different sections of an exam like the Scholastic Aptitude Test.

Key Metrics

Rasch analysis uses the degree of fit to the model to determine *construct validity*. A perfect fit of 1 is unlikely to be achieved with any instrument. So as with FA, researchers have established acceptable ranges. In Rasch terminology, InFit (weighted chi square) should fall between .6 and 1.4. All 12 dimensions of Innovator Mindset fall within this range. Seven of the 12 IM scales fall within a more stringent range between .8 and 1.2 that is recommended for high stakes decisions. By testing for fit, Rasch performs what is analogous to the confirmatory analysis (convergent validity) of FA.

In FA, the eigenvalue of an assessment item is the percent of variability in the characteristic (trait, preference, skill, etc.) being measured, that is explained by that item. Rasch analysis reverses this and calculates an eigenvalue for the percent of variability that is *not* accounted for by the *combined* items that measure that characteristic. So with Rasch, the objective is to achieve a low eigenvalue. For the 12 dimensions of the IM instrument, eigenvalues ranged from 9.1% to 16.7%. So explained variances ranged from 83% to 91%.

Rasch measures two types of reliability, *person reliability* and *item reliability*. Person reliability is an indication of how consistently a test measures what it is supposed to. This is equivalent to *internal consistency* in FA. Person reliabilities ranged from .70 to .88, with all but two of the 12 dimensions above .8 This measure is comparable to the Cronbach's Alpha and KR20 calculations used in FA, with Rasch values tending to come in a bit lower for the same data. So a given Rasch person reliability score is usually comparable to a slightly higher Cronbach's Alpha. Rasch item reliability measures how consistently the same

Post Hoc Analysis - Correlations between IM Scales

Correlations Matrix

	Idea Cognitive *IC*	Idea Values *IV*	Idea Behavioral *IB*	Action Cognitive *AC*	Action Values *AV*	Action Behavioral *AB*	Reality Cognitive *RC*	Reality Values *RV*	Reality Behavioral *RB*	Feedback Cognitive *FC*	Feedback Values *FV*	Feedback Behavioral *FB*	Idea Phase *IP*	Action Phase *AP*	Reality Phase *RP*	Feedback Phase *FP*	Cognitive Profiles *CP*	Values Profiles *VP*	Behavior Profiles *BP*	Innovativeness Index *Index*
IC																				
IV	0.45																			
IB	0.49	0.62																		
AC	0.54	0.52	0.52																	
AV	0.46	0.69	0.59	0.57																
AB	0.48	0.67	0.60	0.63	0.68															
RC	0.28	0.31	0.23	0.42	0.41	0.39														
RV	0.38	0.66	0.59	0.55	0.72	0.55	0.45													
RB	0.29	0.40	0.35	0.49	0.41	0.44	0.57	0.51												
FC	0.39	0.41	0.40	0.52	0.46	0.55	0.59	0.48	0.46											
FV	0.41	0.69	0.52	0.55	0.70	0.69	0.55	0.76	0.60	0.61										
FB	0.35	0.52	0.35	0.48	0.46	0.47	0.50	0.54	0.52	0.45	0.63									
IP	0.77	0.84	0.86	0.64	0.71	0.71	0.33	0.67	0.43	0.49	0.66	0.50								
AP	0.57	0.72	0.66	0.84	0.86	0.90	0.47	0.70	0.51	0.59	0.75	0.54	0.79							
RP	0.39	0.56	0.48	0.60	0.63	0.56	0.82	0.80	0.84	0.62	0.78	0.63	0.58	0.68						
FP	0.45	0.64	0.51	0.61	0.64	0.67	0.65	0.71	0.63	0.81	0.88	0.83	0.65	0.74	0.81					
CP	0.72	0.55	0.53	0.80	0.62	0.67	0.74	0.61	0.59	0.81	0.69	0.58	0.72	0.80	0.79	0.82				
VP	0.48	0.87	0.66	0.62	0.88	0.74	0.48	0.89	0.54	0.56	0.89	0.61	0.82	0.86	0.78	0.81	0.70			
BP	0.53	0.72	0.76	0.69	0.70	0.82	0.54	0.71	0.73	0.60	0.79	0.76	0.82	0.86	0.80	0.85	0.77	0.83		
Index	0.62	0.79	0.71	0.76	0.81	0.81	0.63	0.81	0.67	0.71	0.87	0.70	0.86	0.91	0.86	0.90	0.88	0.93	0.94	

items would perform when given to another similar group of persons. Item reliabilities for the 12 dimensions ranged from .96 to 1.00. These are all well above the .9 threshold that is considered acceptable. There is no FA equivalent for item reliability.

Post Hoc Analysis

A post hoc analysis was done correlating the 12 IM dimensions and their combinations. (See preceding Table.) That analysis confirmed 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
4) Dimensions are most strongly correlated to the Innovativeness Index.

These findings show that Innovator Mindset behaves as one would expect if it is performing as designed.

Differential Item Functioning

Innovator Mindset was analyzed for gender bias using the Mantel test and the Rasch-Welch t-test. No systematic gender bias was found. Nevertheless, given the nature of mindset, it seems likely that it is subject to cultural influences. So until further research can be done, it should be assumed that IM results are somewhat culturally sensitive. This is not to say that IM is biased in a way that reduces its validity or reliability, but rather that where there are differences in innovativeness that may have cultural origins, it will detect them. Theory, data and practice all indicate that anyone has the ability to change their IM score to whatever degree they choose.

External Validation

The external validity of the Innovator Mindset instrument, and the impact of the things it measures, were evaluated by giving this assessment to more than 300 diverse entrepreneurs. Their scores on the Innovativeness Index were compared to the performance of their ventures. This posed some challenges in analyzing that data. Since at least as far back as the seminal work of Nobel Laureate Herbert

Simon in the 1950s, researchers have known that business performance data—things, like revenues and profits—do not fall into a normal statistical distribution. Rather, they typically follow a power law distribution or exponential curve with a few high end outliers creating a disproportionate share of vale. (Think of companies like Google, Facebook and Uber.)

Innovator Mindset scores however do fall into a normal distribution. This mismatch between these this two different types of data distributions means that they cannot be compared using many of the most commonly used statistical tools, such as a t test and least squares (*r* value). So a quasi Beyesian approach was used instead. The Innovativeness Index was segmented into 5 and 10 point quantiles and business performance data was averaged within each of those quantiles. This revealed a dramatic upward trend in value-creation as innovativeness scores increased.

Similar challenges came up during hypothesis testing, when determining the statistical probability that the findings could have been the result of random variability (*p* value). Rather than use calculations designed for normal distributions, the data was put into an Excel spreadsheet and 10,000 simulations were run. The relevant *p* values were determined by the results of those simulations. All were found to be $p < .05$ and in some cases $<.01$

Appropriate Use

Innovator Mindset measures a person's degree of innovativeness. It also provides feedback that someone can use to enhance their innovativeness (and subsequently change their score on this assessment). So IM can be used in two primary ways 1) *Determining Innovativeness*—by measuring how innovative someone may currently be, and 2) *Developing Innovativeness* by identifying adjustments someone can make to become more innovative. This makes IM useful in a wide range of applications.

Determining Innovativeness is something that has value for evaluating potential employees, especially in settings such as R&D, new product development and other roles related to innovation. IM scores have also been found to be remarkably predictive of success and value creation by entrepreneurs, something that should be of interest to investors, who must evaluate candidates seeking funding. When used in these ways, IM should be treated as a diagnostic or source of insights to inform those decisions, but not as the sole criteria for making those selections and

not based on any specific cutoff score. Rather it should be treated as one of many factors considered, and as a way to identify opportunities for personal development in those candidates.

Developing Innovativeness is appropriate in student settings, including college and upper secondary level courses in innovation and entrepreneurship, as well as in startup incubators and business accelerators. Anywhere where aspiring entrepreneurs are developing their personal acumen at launching new ventures or someone is undertaking any innovation initiative. To avoid introducing bias or creating perverse incentives, instructors should not grade students based on their IM scores. Rather, students should be evaluated based on how well they understand how to develop their innovativeness.

Innovator Mindset is useful for raising self-awareness and developing general mental agility, adaptability, resourcefulness, creativity, critical thinking and problem solving capabilities in any context. This includes corporate, academic, non-profit and government. When used to develop innovativeness, the Innovator Mindset assessment provides high value feedback. However to be most effective, it should be part of a broader intervention aimed at teaching the concepts of personal innovativeness and developing the habits of successful innovators. IM also provides a foundation for enlightened leadership development. One that enhances a person's ability to lead innovation and their general leadership skills.

In all of these applications, it is important to recognize that mindset, unlike many personality traits, is dynamic and malleable. IM reports are called *Snapshots* because they capture someone at a particular point in time. Mindset can change based on context and on life experiences, as well as due to an intervention such as coaching or a workshop. This is why a mindset-based approach is such a powerful strategy for personal and professional development.

RESOURCES

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.

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

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.

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

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

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

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

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 Inquiry*. - 1995. - 4 : Vol. 6. - pp. 267-285.

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

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.

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.

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

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.

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

Lazarsfeld PF, Henry NW. Latent structure analysis. Boston: Houghton Mifflin, 1968.

Linacre John M WINSTEPS 3.81.0. - 2014.

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

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

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

Lord FM, Novick MR. Statistical theories of mental test scores. Reading, Massachusetts: Addison-Wesley, 1968.

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

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

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.

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.

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.

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

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

Stauffer, D.A. (2015b), "Evaluating mindset as a means of measuring personal innovativeness", *International Journal of Innovation Science*, Vol. 7 No. 4, pp. 233-248.

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

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.

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

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.

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

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