The Qualitative Report

Volume 18 | Number 41

Article 1

10-14-2013

Prototype Modeling vs. Item Response Theory – A Paradigm Shift for Measurement Professionals

Joseph Yeager
Sommer Consulting Inc.

Linda Sommer
Sommer Consultng Inc.

Juliette Faughnan
Sommer Consulting Inc.

Christophe Geerkens
Sommer Consulting Inc.

Recommended APA Citation

Yeager, J., Sommer, L., Faughnan, J., & Geerkens, C. (2013). Prototype Modeling vs. Item Response Theory – A Paradigm Shift for Measurement Professionals. *The Qualitative Report*, *18*(41), 1-13. Retrieved from http://nsuworks.nova.edu/tqr/vol18/iss41/1

Prototype Modeling vs. Item Response Theory – A Paradigm Shift for Measurement Professionals

Abstract
In science, the usual research mode is to discover cause - and - effect mechanisms operating behind superficial reality. Aviation, for example, uses prototype modeling to develop improved aircraft from one generation to the next. In stark contrast to prototype modeling's search for causal mechanisms, assessment technologies remain dominated by probability strategies — which is why psychological assessment lags behind the hard sciences. A major difference in results exists between inferences produced through item response theory versus the cause - effect prototyping strategies of aerospace and other techno - savvy industries. Parallel to aerospace strategies, systems analysis of language, using prototype modeling strategies, provides an effective assessment technology. This paper shows how the result provides causal mechanisms easily employed for assessing, predicting and changing human performance in many contexts.

Keywords
Prototype Modeling, Item Response Theory, Situational Awareness

The Qualitative Report 2013 Volume 18, Article 81, 1-13
http://www.nova.edu/ssss/QR/QR18/yeager81.pdf

Prototype Modeling vs. Item Response Theory – A Paradigm Shift for Measurement Professionals

Joseph Yeager, Linda Sommer, Juliette Faughnan, and Christophe Geerkens
Sommer Consulting, Inc., Newtown, Pennsylvania, United States

In science, the usual research mode is to discover cause-and-effect mechanisms operating behind superficial reality. Aviation, for example, uses prototype modeling to develop improved aircraft from one generation to the next. In stark contrast to prototype modeling's search for causal mechanisms, assessment technologies remain dominated by probability strategies—which is why psychological assessment lags behind the hard sciences. A major difference in results exists between inferences produced through item response theory versus the cause-effect prototyping strategies of aerospace and other techno-savvy industries. Parallel to aerospace strategies, systems analysis of language, using prototype modeling strategies, provides an effective assessment technology. This paper shows how the result provides causal mechanisms easily employed for assessing, predicting and changing human performance in many contexts. *Keywords: Prototype Modeling, Item Response Theory, Situational Awareness*

Psychological assessment tests, as developed through quantitatively-driven item response strategies, suffer the same fate as Zeno's arrow. Zeno, the ancient Greek, had a problem with his bow and arrow. When released, the arrow traveled half the distance to the target. Then it traveled half of that remaining distance. Then it continued, perpetually, to travel half of that remaining distance, never reaching the target. That accurately describes a great deal of probability-based and inferential assessment research in psychology. It means that psychology frequently does not reach its targets.

Current psychological assessment technologies assume that a person's "true score" on an assessment device cannot be known. This belief also falsely presumes that a validity of 1.0 is impossible to obtain. The limitation this belief places on the assessment field is an artifact of the inferential and probability methods commonly employed. Due to this blindspot, assessment professionals fail to recognize intrinsic language mechanisms. That blindspot holds back more effective measurement science with obsolete inferential methods. In contrast to prevailing beliefs, technology exists for accurate prototyping of language mechanisms with reliabilities and validities of "1.0."

Inferential, statistical or probability strategies, by definition, cannot identify causal mechanisms and provide definitive results. The "Null Hypothesis" that tests the probabilities of statistical significance of theoretical concepts is a prime example of an assessment process without definitive result. The null hypothesis merely informs the researcher that his or her results are probably not accidental. By comparison, investigators of an airline crash would not be satisfied with a mere "probable" cause because probability is not a definitive result in life and death situations.

The null hypothesis is well known as the "workhorse" of psychology (Kreuger, 2001). Yet, the null hypothesis in assessment is, arguably, as outmoded and antique as the original Wright brother's airplane.

> The reliance on inference and the null hypothesis creates a never-ending spiral of questions with few answers and many uncertainties. Had aerospace and

> aviation used a similar inferential approach, human bodies would still be falling out of the sky at the rate of snowflakes in a Rocky Mountain blizzard. (Yeager & Sommer, 2007, p. 562)

This blindspot in the community leads to “who can produce the most elaborate statistics” instead of “who can produce the most effective results.” By servicing the illusory goals inherent to non-causal quantification, the assessment community stagnates. This represents a serious loss to the high stakes involved in measuring behaviors such as personality, attitude, interests, motivation and decision making.

This paper compares two competing technologies found in tests & measurement assessment technology. The dominant strategy is the usual item response theory used to create and develop tests. The newer experimental performance prototype strategies represent a game-changing technology. Used in aviation, computer science, and linguistics, prototyping strategies represent a newer technology within psychological assessment circles: *experimental prototyping*.

Experimental prototyping, as opposed to inferential statistical modeling, has significant advantages. Experimental prototyping uses state of the art technology that represents an advance in test and measurement strategies. Experimental prototyping operates in a manner directly parallel to evolution within aviation and computer science and linguistics. The prototyping approach begins and ends with hard copy, physical reality found in the natural structures of language. That contrasts sharply with the probability approach using inferential statistics and items response theory used in a great deal of current assessment practices. In short, quantitative research currently favors the artifacts of inference inherent in item response theory over the causal relations offered by linguistic approaches; this is an oversight predicated by tradition and a dearth of interdisciplinary rigor.

Prototype Modeling vs. Item Response Theory

Psychological assessment, as measured by quantitatively driven, inferential item response strategies, suffers the fate of Zeno’s arrow in that it doesn’t often hit an effective target. Reliance on probability virtually guarantees that no definitive causes will be identified. By servicing the false god of irrelevant quantification, the assessment community stagnates. The community overly relies on inferential strategies to the detriment of alternative and more effective research methods. In contrast, assessment strategies that utilize language prototype modeling are able to realize 1.0 reliability and validity, as we shall see.

A Flawed Assumption

For many professionals, the claim that 1.0 validity and reliability are even possible in assessment will be difficult to accept. This is because in assessment circles, it is accepted wisdom that validity is a statistical function of a test’s reliability. Classical test theory, moreover, posits that every test taker has a *true score, T,* and that this true score would be the test result if there were no errors in measurement. The assumption, of course, is that a true score cannot be measured—and this is why test reliability is used to determine validity.

These assumptions have produced generations of stagnant *conceptual* measurements due to the failure of item response theory to measure the “real thing.” This kind of logic is unacceptable in the hard, material sciences. The assumption: validity cannot reach 1.0 in reality because behavior can’t be measured precisely. That means that the entire field is held hostage, if not to the logic of Zeno’s Arrow, then, at least, to systemic uncertainty. The assumption prevails that no measurement of behavior is reliable enough to warrant the use of “ratio” measurements.

The fundamental problem here is that estimates are presumed to suffice when measuring elusive concepts of behavior via item response theory. As we have already indicated, one alternative measurement strategy is systems analysis via *linguistic* assessment. Language *is* behavior. Language as a system is not ephemeral or merely conceptual. Language is as “material” in its architecture as is the Periodic Table of the Elements in chemistry. Language systems permit *ratio* measurement as opposed to the less sophisticated, arbitrary scales of nominal, ordinal and interval measurements (Stevens, 1946).

Statistical measures are not the same as measures of causation. R.A. Fisher in 1922 has stated that the psychometric rationale for statistics is: “The object of statistical methods is the reduction of data.”(as quoted in Lehmann, 1998, p. 311) In measuring and managing changes of behavior at the engineering level, certainty of the mechanism of causation is required, not data reduction. The object of behavior in applied settings is controlled, predictable and effective change. No assessment tools designed with item response technology currently offer the medical equivalent of a specific mechanism for diagnosis, prescription, and change of behavior.

For example, explaining to a client-executive that a public speaking phobia is a common experience does not change the phobic behavior. It is well established in behavioral engineering that such a symptom is measured, prescribed and cured with definitive linguistic mechanisms. The same certainty of results has been achieved with linguistic prototyping in a wide range of contexts such as leadership, education, performance enhancement, advertising, competition and organizational development (Yeager, 1983).

Linguistic prototyping and assessment match the design of hard science assessment. The National Aeronautics and Space Administration (NASA) illustrates the design and process required to develop prototypes and finished products in aerospace technology (2011). This NASA developmental process mimics the process used to develop linguistic models. The model-development process serves to create prototypes for behavioral assessment and behavioral change technology.

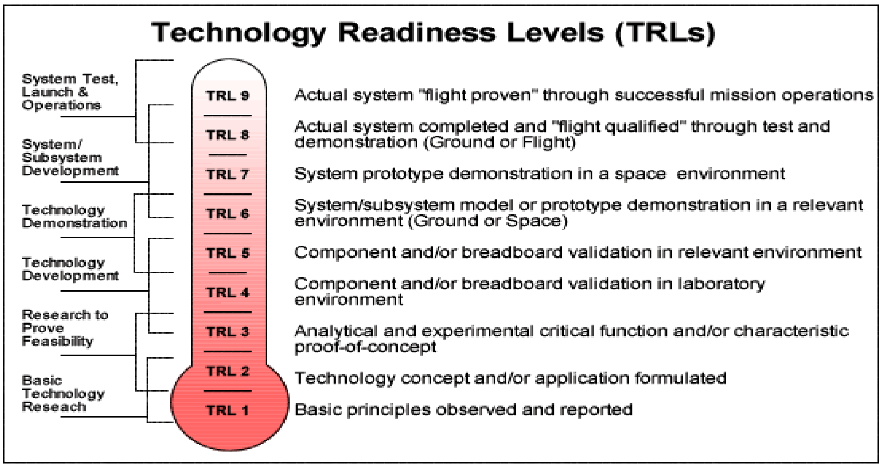

Figure 1. Technology Readiness Levels Retrieved from https://www.spacecomm.nasa.gov/spacecomm/programs/technology/ Reprinted under public domain provision for U.S. government publications.

Systems engineering of language entities requires the identification and inclusion of the whole system and all of its parts. Ineffective “estimation” of its structure and functions guarantees failure. “Estimation” of the necessary flight characteristics of a modern jet or spacecraft guarantees disaster. Who would fly in an aircraft that lacked precision design and proof of its airworthiness? Why would assessment experts rely upon estimates and blue-sky concepts? Answer: No intelligent expert would use inherently flawed metrics. The tests and measurements industry has yet to embrace prototype strategies toward the development of assessment technology.

The Heart of the Matter

Management, as a term, comes from the Latin, *manus* (i.e., the *hand*), with which one manipulates change. An assessment tool based on language prototype modeling measures the way people manage change to their personal advantage as well as to the advantage of the organization that employs their services. An item response strategy deliberately, and ineffectively, separates the assessment questions from the answers. The question and answer are replaced with a correlation coefficient. This is not a useful strategy when decision makers need to manage the affairs of their team members.

For example:

- Question: "At parties I…"
- Answers: a). talk to others, b). keep to myself, c). hide when the boss arrives.
- Result: If the answer is "c). hide when the boss arrives" and that answer is assigned a weighted score of say, 3 points, then the substance of the real answer "hiding from the boss" is replaced with a number, i.e., a "3."

At that point, the assessor has lost the opportunity to know the actual behavior that has been replaced with a meaningless numerical index. The scores of numerous test items then must be aggregated into a trait or a concept score that is labeled and must then be "interpreted" for its inferential meaning to be estimated. To wit:

- Conclusion: This is not an optimal way to assess or manage behavior.
- Corollary: Measuring vague concepts is counterproductive.

That example shows how abdicating control over the Q&A process is not an optimal strategy. At the point of separating the Question from the Answer, control over the data and the mechanism of action of the behavior is lost. The Q&A process as illustrated here takes hard data and reduces it to ordinal or interval data. In prototyping, the researcher requires ratio data or the machinery of the decision at hand will be irretrievably lost. At that point in time the would-be assessors have launched the assessment equivalent of the fickle and unmanageable flight common to a hot air balloon. The practical goal is to have a 1 to 1 correspondence between cause and effect (i.e., between stimulus and response). In aviation, in behavioral interventions, and in prototyping strategies the 1 to 1 goal is a requirement.

With linguistics as the basis of a prototype strategy, when assessors ask an assessment question (e.g., in questionnaire format or in interviews) and can process the answer from the stimulus to its outcome-effect, *they have a controlled strategy*.

For example:

- Question: "What do you do at parties?"
- Answer: "I like to avoid the boss because I don't want to take the chance of making a bad impression."
- Result: Now you have gained from the Q&A process, at the very least, a goal, a worrisome possible outcome, and information about a lack of rapport skills.

- Conclusion: When assessors capture that kind of data and its inherent linguistic mechanism of action (using linguistic prototyping strategies), they have the equivalent of a modern jet aircraft that is under their control. That control will produce the desired outcome of effective assessment and management of performance.

The heart of the matter regarding correlated item response theory and controlled experimental prototypes comes down to whether opinion reigns or whether one has a 1 to 1 causal correspondence among the phenomena at hand. The difference in control is between rolling dice and getting a guaranteed dividend. Item analysis' quantification strategy can never reach a validity of 1. However, in the real world of product designs, like aircraft or computer programs, 1 to 1 validity is required or failure is a certainty. Language based experimental prototyping uses the same essential approach as NASA, as illustrated in Table 1. Here is a comparative breakdown:

Table 1. Experimental Prototype Strategy versus Item response Strategy

Controlled Experimental Prototypes	**Correlated Item response Theory**
Requires Mechanism of Action (MOA)	Assumes Correlations without MOA
Direct Observation of Behavior (Actual)	Hypothesize a Construct (Theoretical)
Identify Behavioral Whole and Parts	Assumptions & Behavioral Theory
Identify an Applied Product Need	Identify a Theoretical Gap
Build a Test Model of the Behavior	Construct a Questionnaire of a Concept
Test the Prototype in Real Environments	Create Comparable Questionnaires
Determine Performance Characteristics	Compare the Concept Against Itself
Build & Fully Test Operational Prototypes	Refine the Concept Against Criteria
Begin Production of the Test Product	Begin Production of the Test Product
Validity is a Function of Real Performance	Validity a Function of Correlated Estimates

Item Response Theory: A Representative Example

Language clearly drives behavior. Experimental modeling of language behavior yields excellent results when measuring high-stakes job behavior. The idea that language can produce behaviorally relevant data is at least as old as Sir Frances Galton's lexical hypothesis. Galton proposed that life experience becomes encoded as language. Building experimental language prototypes does make a significant difference. For instance, famed psychologist Raymond Cattell developed the highly regarded 16 Personality Factor test using an early linguistic approach. His work developed a few core concepts by using factor analysis.

In essence, he boiled down the thousands of personality-related adjectives in the dictionary to the well-known 16 factors. But the now-obsolete factor analysis "hot-air-balloon" strategy suffers from the item-analysis-related absence of control. The strategy lost the crucially important mechanism of action inside the decision's language process, structure and content.

Another representative flaw in that item response strategy occurs in the first few items of the test's results. The first few patterns that "explain" most of the variance are warmth, reasoning, emotional stability and dominance. Those static patterns are *always* weighted to matter a great deal to assessors. That is akin to saying that the lungs are the most important part of the body because they process oxygen. In a system like a car or a decision, it is not useful to say the engine, or any component, is the most heavily weighted part of any complex system. In part, importance depends on behavioral context. And, in part, a system requires that all the parts operate as a system or the system fails.

Ideals, Concepts, and Fudge Factors

That “most important” implies an *ideal* component in the system akin to the status assigned to a high IQ. In business, a common context for assessment, “ideal” versus what works are usually quite different things. In practice, every part is important because it is a system. An ideal is a static notion but organizations are dynamic in their adaptation to competition. The built-in bias of *ideal* as applied to profile information is not specifically useful in assessment situations. Often there is a false assumption of a single correct value or response. In the domain of decision making, most any pattern can change its role and relevance within a decision strategy. Change depends a great deal on the person’s perception of the shifting contextual demands.

A carpenter or test-builder who builds us a *conceptual* or *inferential* assessment door from ordinary item analysis will leave us out in the cold when winter arrives. Many obsolete tests and measurement technologies leave us in the cold. The fact that obsolete technologies are replicated on computers does not change their obsolescence. As computer scientists say, GIGO (i.e., garbage in, garbage out).

In contrast, the last few decades of experimental prototype modeling of language behavior has accomplished major advances in assessing and predicting performance. (Yeager, 2003) Instead of static assessment portraits, assessors want the priorities in actual decision making to change to meet the situational demands of a job’s role. By analogy, when driving, you step harder on the gas to go uphill than when on level terrain. Pavlov pointed us in the general direction with his prototyping of behavior through experimental modeling. Modern linguistics has gone much further and does make a significant difference in the gains that assessors can expect.

One thing Pavlov did not know was that language produces a unique phenomenon called “one-trial learning.” That is, stored memories can be recalled and put into use immediately without repeated memorization of new behaviors. Suppose you, the reader, were convinced by some means that the page you are holding is radioactive. You would not need to be instructed to drop the dangerous page. That is an example of “one trial learning.” In linguistic prototyping strategies, such a phenomenon has enormous capability to rapidly assess, modify and develop performance related behaviors.

The bulk of item response theory researchers can never reach their targets because their method makes it impossible. Any validity of less than 1.0 is unacceptable in the hard sciences that undergird aerospace and computer science. The fields of aerospace and computer science stand in contrast to this assessment rationale. Aerospace and computer science require causal relationships among the researcher’s tool kit as well as the results. In aviation, a method that could never reach “1.0” validity means that no aircraft would ever be able to fly for sure, only probably. Who would ride in such a craft?

This presumption leads to another ancient Greek of note, Procrustes. Procrustes’ guest bed only accepted one size of guest. If people were too long, they were cut to fit. If too short, they were stretched to fit the bed. Psychology’s inferential methods have become the Procrustean bed of assessment tools. The methods predetermine the kind of results that will be achieved. When test items are dropped out of a population of test items, the statistical answer to this flaw is to employ a statistical correction (i.e., a “fudge factor”).

A well-known example of a fudge factor was Einstein’s addition to his theory of relativity in an attempt to justify his assumption that the universe was stable and not expanding. That aspect of his theory was wrong. But it was "only theoretical." Can anyone imagine that happening in aeronautical engineering?

Systems Perspective: Language Prototype Modeling

A major upside of language analytics, as applied to behavior, is its portrayal of motives and decision making as a whole, integrated system, not as a collage of fragmentary data. This system is the system of 'meaning making' known as language. As stated in grade school grammar books: A sentence is a group of words expressing a complete thought. A complete thought, then, is a whole linguistic and behavioral entity that can be parsed for its structure and its components. To understand Humpty Dumpty's motives and decisions, it is not especially useful to smash him into fragments with questioning techniques that obtain bits and pieces of him instead of directly and systematically observing and asking about what makes him tick – in whole and in part. Observation of Humpty's linguistic mechanisms, i.e., linguistic content, structure and linguistic process, will reveal the desired answers. Let us look at how this works.

Similar to Tip O'Neill's coined phrase "All politics is local," all decisions are contextually local to the individual making the decisions. However, all decisions use the same essential linguistic technology. These two generalizations about linguistic technology serve as effective base-line assumptions in prototyping linguistic behaviors.
That is:

- Decisions are complete systems with identifiable linguistic and behavioral parts.
- All decisions are immediately and contextually local to the decision maker.
- All decisions operate with linguistic and behavioral mechanisms.

Whether writing ad copy for a luxury automobile, teaching a program, writing a term paper or giving a campaign speech, an individual uses the same basic language repertoire when making a decision. Language is a large system of coded, identifiable elements that operate as a whole system. Language analytics provides linguistic researchers and assessment professionals with scientific tools to measure and predict outcomes. As one example:

> Linguistics' unique ability to analyze and score the language of the target audience provides insights beyond those of typical research. The results are immediately actionable for marketing, advertising, or sales interventions. The main principle: match the message to the content and linguistic structure of the audience mind-set.
>
> In research directed at new product design, physicians were asked, "What do you want in an ophthalmic suture?" The prototypical answer was: I don't think there are any new breakthroughs in needle technology. I want improvements so I can do a running stitch without the thread tangling and I want the needle to stay sharp so it pierces tissue nicely and I don't want it to magnetize and stick to the tweezers. (Yeager, 2003, p. 147)

The expressed thoughts of those physicians are an example of a minor-sized linguistic system. Again, this reflects the reality that any sentence is defined in grade-school grammars as a group of words expressing a complete thought. In other words, a sentence is a system. A sentence is also a sub-system in a wider discourse such as a paragraph, a conversation or a chapter in a book. When a written or verbal communication is framed and modeled as a prototype system, there are many implicit structural variables (Stogdill, 1970; Freidenthal et al., 2012). Such prototype variables are found in a linguistic response like that of the above physician example, even when excluding elements such as standard English grammar.

Language Features	**Transformational Grammar**
Parts of Speech	Decision System Processing
Subject Verb Object	Linguistic System Hierarchies
Figures of Speech	Decision Frames & Context
Dialects	Semantics & Syntax
Sentence Parsing	Meta Patterns
Diction	Matching Strategy
Syntax	Analog Structure & Strategy
Idioms	Emotional & Verbal Synesthesia
Validity is a Direct Performance Measure	Validity is an Inherent Language Function

Table 2. Language Features Compared to Transformational Grammar

Explicit Strategies

Linguistic systems analysis strategies are arguably outgrowths of qualitative strategies. Certainly the long history of content analysis superficially resembles the kinds of data generated by linguistic strategies. For the sake of illustration and familiar points of reference to readers, that point of view will be used in this paper. But please keep in mind that there are mechanisms involved in linguistic systems technology that are not enjoyed by quantitative nor common qualitative strategies.

Traditional statistical tools permeate the professional contexts of most behavioral experts. Those professional preferences for statistical tools then frame how those experts select their strategies for studying motivation. The choices then made by those same experts often rely on an implicit scorecard of statistical "frames" that bias experts to favor a particular approach. As Kruger (2001) noted above, the statistical approach has been dominant.

Reflexive professional preferences for statistical approaches act like the lost nail in Ben Franklin's metaphor. "For want of a nail the horseshoe was lost; for want of the horseshoe the horse was lost; for want of a horse the rider was lost; for want of a rider the battle was lost." This paper wishes to make some of the differences among strategies explicit in order so that experts can make more effective choices. More winning battles for progress in behavioral science might result.

The goal here is to examine the respective roles and defining features of measurement frames used in advanced motivational profiling, ranging from individuals to large populations. The essential frames of reference are shown in Table 1.

Examples of linguistic variables in a prototype model include: linguistic frames, belief frames, change tense, semantic space, directional tense, and unconscious cognitive processing. In the physician sample, coding decisions in terms of verbal and written messages gave the needed answers. The corresponding message design elements of those variables produced sales increases of that ophthalmic product by more than 300% in less than 90 days. The decisions of many audiences, customers or employees, for example, are directly impacted by linguistic interventions (Yeager, 2003). The decisions of executives, job candidates and employees are equally as accessible to similar linguistic assessments.

Casual reading of The Wall Street Journal, Business Week or Fortune magazines, for example, provides instant confirmation of the outstanding role of decision-making. While always an important informal element in interviews, linguistic assessment strategies have been incorporated into the prototypes and design of psycholinguistic test development (www.SommerYeager.com).

Scientific Mechanisms of Action

These two strategies (item response and linguistic) demonstrate major differences in the assessor's control over the assessment process and its outcomes. The researcher's control within item response practices can be compared to the unpredictable journeys of a hot air balloon. Flying in hot air balloons is a process of estimation, not precision control. In contrast, experimental prototyping strategies can be compared to the precision flight controls of modern jet aircraft.

In a balloon, as with common item response strategies, assessors know where they've been, but they don't know quite where they are going with their correlations. Balloonists and psychological assessors have no control over the trip, and no control over how everything will land (and there is no going back the way one arrived). In the long run, tests and measurement practices seek scientific mechanisms of action. The reward is the kind of control found in competing hard sciences such as found in aerospace professions.

In general, many popular test development processes offer the developer and assessor little real control over the journey to the desired destination. Typical statements found in the conclusion sections of professional literature usually offer us *careful interpretations* and *suggestions for further research*. Simply put, customary test development procedures offer only equivocal promises. They do not *show me the money!*

It is fair to say that item response strategies produce an outcome that is explained in large part by the "Forer Effect" (Forer, 1949). In 1948, psychologist Bertram R. Forer gave a personality test to his students. He told his students they were each receiving a unique personality analysis that was based on the test's results and to rate their analysis on a scale of 0 (very poor) to 5 (excellent) on how well it applied to themselves. In reality, each received the same analysis. In essence, Forer presented these characterizations:

> You have a great need for other people to like and admire you. You have a tendency to be critical of yourself. You have a great deal of unused capacity which you have not turned to your advantage. While you have some personality weaknesses, you are generally able to compensate for them. Disciplined and self-controlled outside, you tend to be worrisome and insecure inside. At times you have serious doubts as to whether you have made the right decision or done the right thing. You prefer a certain amount of change and variety and become dissatisfied when hemmed in by restrictions and limitations. You pride yourself as an independent thinker and do not accept others' statements without satisfactory proof. You have found it unwise to be too frank in revealing yourself to others. At times you are extroverted, affable, and sociable, while at other times you are introverted, wary, reserved. Some of your aspirations tend to be pretty unrealistic. Security is one of your major goals in life. (Forer, 1949, p. 120)

On average, the rating was 4.26. But only after the ratings were turned in was it revealed that each student had received identical copies assembled by Forer from various horoscopes. As can be seen from the profile, there are a number of statements that could apply equally to anyone. These statements later became known as Barnum statements, after showman P.T. Barnum.

Proponents of item response theory point to regularities in research findings such as the “Big Five” personality traits theory. However, finding regularities does not imply causation. In science, cause and effect are required. Regularities, we must point out, are found in astrology, the Forer effect, and many superstitions. Regularities of results can have only speculative or theoretical connections to cause-effect results. “Magic” or other beliefs about non-physical causality does not supply the actionable causality required in science.

Walter Mischel (2008), past president of the American Psychological Society, has introduced the profession to awareness of its “toothbrush problem.” As Mischel notes, no newly minted, self-respecting psychologist would use another psychologist’s theory any more than he would use the other person’s toothbrush. The “toothbrush problem,” as reluctantly recognized by psychologists, shows an emerging awareness of the fundamental problems of ineffective approaches. In contrast, prototype strategies produce robust cause-effect results that are not artifacts of the methodology being used.

In the real worlds of business, aviation and justice, the differences in consequences between typical item response strategies versus the experimental prototype strategy are, literally, life and death matters. The consequences and effectiveness of applied experimental prototyping practices offer much more effective results than test & measurement assessment strategies. Part of the answer lies in the simple need for more rigorous cause-effect control over the whole process which is, by definition, impossible with an item response strategy in test development.

Luddite Case Example

As with any technology, such as the actual contributions of experimental modeling of decision making, one can expect resistance from the “not invented here” people. Many have careers invested in the obsolete technology, as did the Luddites of old England. Luddite-like attitudes can be found in any field. The term Luddite refers to people who are against any form of modernization, including technology. These attitudes will always be with us and accompany all forward looking innovation.

Yeager and Sommer (2005) document a recent example of Luddite-like attitudes. An author writes the lead article in the flagship journal Psychological Science. In the paper the author describes how he used an item response strategy in an attempt to modify the policy and behavior of executives at the National Institutes of Health. The author bravely confessed to bewilderment at the NIH rejection of his selection strategy.

The NIH executives were tasked with creating a “fair” selection process in making research grants. The researcher’s earnest work was rejected by the executives because the answers were of no use. This case illustrates the role of “frame blindness.”

> A researcher framed the task of refining a ratings process as a research project when he could have done better had he framed it as a qualitative consulting project. He needed to persuade the clientele. Instead, he framed his work as a justification supplied by research literature itself. The [client] executives, for at least one very good reason, rejected the result. To repeat their crucial point, “The scientific data aren’t relevant.” It is debatable whether this quantitative approach was actually scientific. It was quantitative and statistical, but scientific? Not necessarily.” (pp. 492-495)

This researcher’s work represents a great many current selection strategies.

Change can be harsh. In business and most organizational behavior, results do matter. Experimental linguistic prototyping has produced a highly effective alternative to “item response theory” and its related applications. To illustrate the occasional resistance to innovation, we can refer to how the leading scientists of the world reacted to Einstein’s publication of his “Theory of Relativity.” The critics banded together and wrote a counter-argument to Einstein’s theory, and it was presented in a paper called “100 Against Einstein.”

When a reporter asked Einstein what he thought about their rebuttal, Einstein replied calmly: “If they were right, it would only take one.” So, we can comfortably reject the bulk of tests and measurement critics who might object to experimental prototype strategies as a modern alternative to probabilistic approaches. The limits of the inferential approach to assessment are made clear by experimental linguistic prototyping practices.

Summary: Situational Awareness and Blindspots

Situational awareness involves being aware of what is happening in the vicinity. The purpose in psychological assessment is to understand how information, events, and one's own actions will impact goals and objectives, both immediately and in the near future. When an entire field of study is blind to the circular effects of probabilistic methodology, situational awareness is lacking. Without an accurate perception of the overall field of play, i.e., cause-effect science, then opinions will reign where technology is lacking. Mere quantification is not an adequate solution.

Traditional assessment psychology has exploited methodologies that limit forward movement. Examples of areas that use traditional methods are found among tests of personality, attitude, aptitude, decision making, motivation and beliefs. Popular tests routinely employ methods that cannot produce definitive answers. The tests and measurements field has settled for approaching problems within limits that produce equivocal answers. Equivocal answers produce a public impression of ineffectiveness about the profession.

The fact that traditional methods cast a long shadow on the assessment field tells us how late in the day it is. Linguistic prototyping strategies have advanced test construction technology. This technology produces new levels of effectiveness in comparison to the traditional practices of item response theory. Behavioral assessment, especially compared to more competitive and effective sciences, could use a reassessment of blind psychometric assumptions. The rewards found in such a reassessment can move the field forward to the benefit of all concerned.

References

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society. Series A. 222,* 309-368.

Forer, B. R. (1949). The fallacy of personal validation: a classroom demonstration of gullibility. *The Journal of Abnormal and Social Psychology, 44*(1), 118-123.

Friedenthal, S., Moore, A., & Steiner, R. (2012). A Practical guide to SysML (2nd ed.). *The Systems Modeling Language.* Waltham, MA: The Morgan Kaufman MK/OMG Press.

Krueger, J. (2001). Null hypothesis significance testing: On the survival of a flawed method. *American Psychologist*, *56*(1), 16-26. Retrieved from http://www.psycinfo.com/plweb-cgi/padirect_icnstf.pl

Lehmann, E. L., & Casella, G. (1998). *Theory of point estimation*. New York, NY: Springer.

NASA. (2011). Technology. Retrieved from https://www.spacecomm.nasa.gov/spacecomm/programs/technology/default.cfm

Stogdill, R. M. (1970). *The process of model-building in the behavioral sciences*. New York, NY: Norton.
Stevens, S. S. (1946). On the theory of scales of measurement. *Science. 103*(2684), 677–680. doi:10.1126/science.103.2684.677. PMID 17750512. Retrieved from http://web.duke.edu/philosophy/bio/Papers/Stevens_Measurement.pdf. Retrieved 16 September 2010.
Yeager, J. (1983). A model for executive performance coaching. In J. S. Manuso (Ed.), *Occupational clinical psychology* (pp.129-146). New York, NY: Praeger Scientific.
Yeager, J. (2003). Innovative motivational profiling: Comparing marketing projective techniques versus linguistic forensic techniques. *The Qualitative Report, 8*(1). Retrieved from http://www.nova.edu///-1/.html
Yeager, J., & Sommer, L. (2005). How linguistic frames affect motivational profiles and the roles of quantitative versus qualitative research strategies. *The Qualitative Report, 10*(3), 463-511. Retrieved from http://www.nova.edu///-3/.pdf
Yeager, J., & Sommer, L. (2007). Linguistic research strategies versus quantitative research strategies – Different roles, different results. *The Qualitative Report*, *12*(4), 561-579. Retrieved from http://www.nova.edu/ssss/QR/QR12-4/yeager.pdf

Author Note

Joseph Yeager, Ph.D., is a licensed psychologist and is considered by many to be the world's leading expert on performance, persuasion and decision making using rapid-impact methodologies on the scale of individuals, groups and mass audiences. His experience in aviation lead to use of scientific prototyping strategies for developing business applications of decision making technology in M&A, marketing, advertising and human resources. It is the same robust strategy used to advance each generation of aircraft and aerospace developments.

Linda Sommer has expanded the role of communications technology applications into a highly visible force within the international business and professional communities garnering a strong following among the Fortune 100, is a Diplomate of ACFEi and DABLEE and DABPS. She has coached executives and officers in Fortune 500 companies for over fifteen years and is sought after as a keynote motivational speaker and corporate turnaround expert.

Juliette Faughnan, President of Sommer Consulting, brings her decade-plus expertise in psychology and linguistics to bear with a total systems analysis approach. Juliette manages global projects, research innovations, and technology implementation for a wide range of client organizations. Her expertise in behavior and language commands a lead operational role in developing strategic approaches for her international clientele. Juliette is a Magna cum Laude graduate of Duke University where she graduated Phi Beta Kappa. She is currently completing her Masters from the University of Pennsylvania in Organizational Dynamics.

Christophe Geerkens, Sommer Consulting's European Partner, leads his field in innovative decision analysis technology. His expertise includes communications, decision modeling and behavioral modeling of consumer and employee populations. His assessment and decision strategy profiling produce state of the art communications, marketing research and human resource management programs.

Article Citation

Yeager, J., Sommer, L., Faughnan, J., & Geerkens, C. (2013). Prototype modeling vs. Item response theory – A paradigm shift for measurement professionals. *The Qualitative Report, 18*(81), 1-13. Retrieved from http://www.nova.edu/ssss/QR/QR18/yeager81.pdf