Bateson and Wright on Number and Quantity: How to Not Separate Thinking from Its Relational Context

William P. Fisher Jr. [1,2]

1   BEAR Center, Graduate School of Education, University of California, Berkeley, CA 94720, USA; wfisher@berkeley.edu
2   Research Institute of Sweden, Lindholmspiren 7 A, 417 56 Gothenburg, Sweden

Abstract: As part of his explication of the epistemological error made in separating thinking from its ecological context, Bateson distinguished counts from measurements. With no reference to Bateson, the measurement theory and practice of Benjamin Wright also recognizes that number and quantity are different logical types. Describing the confusion of counts and measures as schizophrenic, like Bateson, Wright, a physicist and certified psychoanalyst, showed mathematically that convergent stochastic processes informing counts are predictable in ways that facilitate methodical measurements. Wright's methods experimentally evaluate the complex symmetries of nonlinear and stochastic numeric patterns as a basis for estimating interval quantities. These methods also retain connections with locally situated concrete expressions, mediating the data display by contextualizing it in relation to the abstractly communicable and navigable quantitative unit and its uncertainty. Decades of successful use of Wright's methods in research and practice are augmented in recent collaborations of metrology engineers and psychometricians who are systematically distinguishing numeric counts from measured quantities in new classes of knowledge infrastructure. Situating Wright's work in the context of Bateson's ideas may be useful for infrastructuring new political, economic, and scientific outcomes.

Keywords: epistemology; logical types; number; quantity; measurement; metrology; networks; stochastic pattern recognition

1. Multilevel Meanings and Logical Types

Difficulties in communication posed by the simultaneous presence of multiple levels of meaning in language have been a perennial topic of investigation for philosophers, logicians, anthropologists, and, more recently, theorists in the areas of complex adaptive systems, autopoiesis, organizational research, and knowledge infrastructure. The odd capacity for statements about statements to be meaningless even when they are grammatically correct (as in the viciously circular, "This statement is false") led Russell [1,2] to develop a theory of logical types as a means of distinguishing levels of propositions. Bateson [3] cites Russell along with Wittgenstein, Carnap, Whorf, and his own early work as examples of the need to avoid confusion by separating and balancing concrete denotative statements about observations, abstract metalinguistic statements about words, and formal metacommunicative statements about statements.

In their account of these sequences of nested symmetries, Russell and others assume a Cartesian, dualist perspective of externally imposed control as the only viable means of respecting logical type structures. Bateson and others, however, take an agent-based perspective that lends itself to an alternative internally derived capacity for devising infrastructural standards akin to those of natural language. The utility and meaningfulness of these standards and their adaptive capacities in the face of otherwise unmanageable complexity may provide otherwise unattainable motivations for investing in their construction.

Ecologically speaking, Bateson contends, a fundamental epistemological error is made when cognitive operations are separated from the structural relationships in which they

are immanent (i.e., when the symmetry of adjacent levels of complexity is ignored). Thus, referring denotatively to something learned ("the cat is on the mat") is very different from a metalinguistic statement about what was learned ("the word 'cat' cannot scratch"). Metacommunicative statements about statements, in turn, assert theories about learning ("my telling you where to find your cat was friendly"). Paradoxical sentences ("This statement is false") present what seems to be a factual denotative observation but are simultaneously metacommunicative, because they refer back to themselves in a statement about a statement. Symmetry is violated despite the grammatically correct syntax.

Cross-level confusions such as these are ubiquitous in existing knowledge infrastructures. They can even be said to form a fundamental component in the modern cultural worldview. They have been incorporated unintentionally, as a result of seeming to have no obvious alternative, into the background assumptions, policies, and procedures of contemporary market, governmental, educational, healthcare, and research institutions. These mixed messages continue unabated, in general, even though their root causes have been successfully addressed for decades in specific instances of theory and practice.

Important aspects of these solutions were anticipated by Bateson but came to fruition in the work of Benjamin Wright, a physicist and psychoanalyst who made a number of fundamental contributions in measurement modelling, experimental approaches to instrument calibration, parameter estimation, fit analysis, software development, professional development, and applications across multiple fields [4,5]. After describing the problem of how confused logical types are implicated in mixed messages, Bateson's distinction between numeric counts and quantitative measures will be shown to set the stage for Wright's expansion on that theme. Specifically, the problem being addressed concerns the epistemological confusion of numeric counts and measured quantities, and the solution offered consists of improved measurement implemented via metrological infrastructures functioning at multiple levels of complexity. The specific contribution made by this article is situating measurements and instruments calibrated to reference standards in the context of larger epistemological issues; this connection opens up new possibilities for meeting the highly challenging technical demand for new forms of analysis and information technology that are as locally grounded as they are temporally and spatially mobile and abstract [6–8].

2. The Schizophrenia of Symmetry Violations Ignoring Logical Types

Cross-level fallacies falsely generalize parts to wholes, or vice versa [9]. When wholes are greater than the sums of their parts, and exist at a higher order level of complexity, valid generalizations must be based in persistent isomorphisms demonstrably exhibiting shared functional relationships [10]. Ignoring cross-level fallacies results in a kind of schizophrenia, where messages confusing two levels of meaning put the listener in an unresolvable double bind.

Concurring with Bateson and generalizing his double bind concept to the pervasive incorporation of cross-level fallacies in communications, Deleuze and Guattari [11] suggest that " . . . society is schizophrenizing at the level of its infrastructure" (p. 361). Brown [12] diagnoses the problem as one in which "schizophrenic literalism equates symbol and original object so as to retain the original object, to avoid object-loss." Star and Ruhleder [6] similarly point out that cross-level fallacies' "discontinuities have the same conceptual importance for the relationship between information infrastructure and organizational transformation that Bateson's work on the double bind had for the psychology of schizophrenia."

Repeated experiences with cross-level fallacies in psychological research during the 1950s and 1960s played a central role in Benjamin D. Wright's (1926–2015) insights into how to avoid them. Wright [4,5] was a certified psychoanalyst [13,14] and former physicist [15] whose experience with computers and mathematical modelling made him an early leader in educational and psychological research [16,17]. Significantly, that work involved children with schizophrenia [18,19].

Working with a UNIVAC I, Wright [16] wrote a combined factor analysis and regression program. In his role as a consultant to Chicago-area marketing firms, Wright was "doing 10 to 20 factor analyses a week," with some clients that "did the same study every couple of weeks for years," running "one study a hundred times with the same instruments." Doing this work, Wright [20] said he was in "considerable distress," feeling like a "con man" and "a crook." As a physicist, he sought reproducible measures read from instruments calibrated in interval units traceable to consensus standards. As a psychoanalyst, Wright was concerned with human meaning. Fearing that he was satisfying neither his scientific nor his human values, Wright [20] joked about "going to the psychoanalyst to have my schizophrenia mended week by week."

Bateson's [21] sense of the schizophrenic double bind involves situations in which an implacable authority demands rigid conformity with idealized structures that cannot be realized within the existing constraints of a given situation. This is the kind of situation in which Wright found himself, and that situation remains today the context in which those employing commonly used statistical methods in psychology and the social sciences still find themselves.

The double bind is one in which the value of mathematical abstraction is asserted and desired; at the same time, the methods employed for realizing that value are completely inadequate to the task. Everyone is well aware of the truth of the point made by both Bateson and Wright, that numeric counts are not measured quantities: it is plain and obvious that someone holding ten rocks may not possess as much rock mass as another person holding one rock. The same principle plainly applies in the context of data from tests and surveys, such that counts of correct or agreeable answers to easy questions are not quantitatively equivalent to identical counts of correct or agreeable answers to difficult questions. The practicality, convenience, and scientific rigor of measurement models quantifying constructs in ways facilitating the equating of easy and hard tests have been documented and validated for decades, although have not yet been integrated into mainstream conceptions of measurement.

Even so, the resolution of the double bind Wright helped to create is widely applied in dozens of fields and thousands of papers published over the last 50 years (as can be readily established via searches of Google Scholar for "Rasch" in association with any of the following: analysis, measurement, scaling, model). This resolution has not, however, been incorporated into mainstream methods or the background assumptions and infrastructures of research and practice.

Much of the discord in today's political, economic, organizational, and environmental world exhibits the characteristics of a schizophrenic double bind [11,22–24]. Practical difficulties in formulating clear theoretical and methodological applications of the double bind concept in psychological counseling [25–30] do not remove the systemic need for resolution of the cross-level fallacies embedded within knowledge infrastructures. Star [31] frames the problem, asking,

> In the absence of a central authority or standardized protocol, how is robustness of findings (and decision making) achieved? The answer from the scientific community is complex and twofold: they create objects that are both plastic and coherent through a collective course of action.

In this context, the question arises as to the general feasibility of a systematic experimental solution to the problem of pervasive cross-level fallacies in communication. What collective courses of action resolve double binds in communication by creating objects that adapt to local circumstances without compromising their coherent comparability?

3. Separating and Balancing Levels of Reference

Ever since the publication of Russell's 1908 article on logical types [1], writers working in a wide range of different fields have recognized the need to closely attend to the distinctions among them. In the context of living systems, "every discussion should begin with an identification of the level of reference, and the discourse should not change to another level

without a specific statement that this is occurring" [32]. When developing information technology, "for systems that provide electronic support for computer-supported cooperative work, only those applications which simultaneously take into account both the formal, computational level and the informal, workplace/cultural level are successful" [6]. Therefore, "Multilevel thinking, grounded in historical and spatiotemporal context, is thus a necessity, not an option" [33] (p. 355).

In support of multilevel thinking, Van de Vijver and colleagues [34] offer advanced multilevel analytic methods, Palmgren [35] extends Russell's [1] results on the mathematical logic of a functional reducibility axiom, whereas Lee and colleagues [36] describe the conceptualization of psychological constructs in a multilevel approach to individual creativity. The availability of tools and concepts such as these for making and sustaining cross-level distinctions has not, however, led to the development of broad-scale infrastructures incorporating those distinctions.

The difficulty is one of perceiving and taking hold of a fundamental but implicit characteristic of the social environments in which we exist. Bateson [3] and others [37] note that contextual frame shifts across levels of complexity are not usually made explicit in everyday discourse. Common usage simply moves words across levels without attending to the implicit shifts in meaning across frames. Sometimes funny jokes and meaningful metaphors can be made from these kinds of dissonances, but far too often the disconnection leads only to miscommunications. Adopting Bateson's [3] terms, the question taken up here is how we might devise infrastructural standards distinguishing between denotative statements about learning something, metalinguistic statements of learning about learning, and metacommunicative theories of learning, where statements are about statements.

For a science of complexity or a science of polycontextual metapatterns based on Bateson's work [38–40] to realize its mission, knowledge infrastructures must operationalize metalinguistic awareness. Promising preliminary efforts in this direction [7,8,41–50] overlook an essential set of opportunities and challenges of decisive importance.

Krippendorff [51] provides an important hint in a new direction when he identifies an unresolved issue in Bateson's use of Russell's theory of types. Krippendorff suggests that the omission of self-determination and autonomy from Bateson's sense of logical type distinctions demands the development of a new, more constructive and intentional, "theory that allows circularities to enter and that can then explain, among other things, why belief in the Theory of Logical Types reifies itself in all kinds of hierarchies whose experiential consequence almost always is oppression."

4. Numeric Counts vs. Quantitative Measures

A theory of that kind, one that includes self-determination and autonomy, follows from recognizing and leveraging patterns in the way things come into words. Two primary patterns are of interest. The older pattern dates to the origins of language and involves words associated with semantically well-formed conceptual determinations, where a class of things in the world (cats, for instance) with consistently identifiable features has a name easily learned by all speakers of the language. These word–thing–concept assemblages, or semiotic triangles, have a generality and universality to them that makes habitual usage and transparent representations possible. Usage typically distinguishes levels of meaning appropriately, such that double bind confusions and cross-level fallacies may be found in humorous jokes, poetic metaphors, and deceitful prevarications, but are not deliberately distributed throughout messaging systems.

A more recently emerged pattern, however, involves words associated with things that are neither consistently identifiable in the world nor clearly determined concepts. This class of words often involves numbers treated as meaningful in the same way that the word "cat" is, and in the same way that the number words for meters and hours are, but which plainly and obviously cannot and do not actually support that kind of generality. This class of words concerns scores from educational tests and ratings from assessments and surveys. In classrooms, hospitals, workplaces, the media, and elsewhere,

counts of correct answers, sums of ratings, and associated percentages are treated as meaningful, even though their signification is inherently restricted to the context defined by the particular questions answered. Common usage incorrectly assumes that the denotative factuality of numeric counts and scores translates automatically into quantitative measures. Denotatively meaningful counts of correct test answers are fallaciously applied at the metalinguistic level as representations of performance. They are similarly also used at the metacommunicative level to justify instructional, admissions, and graduation decisions, even in the absence of metalinguistic significance and an experimentally substantiated theoretical rationale.

The cross-level transference of scores whose meaning depends strictly on the specific content of particular questions asked of particular people results in the kinds of vicious circularities and schizophrenic double binds reported by Wright [20] concerning his use of factor and regression analyses of test scores. Problems associated with this kind of cross-level fallacy also have been observed in recent years as a consequence of the No Child Left Behind education act in the United States. [52]. These problems can potentially be rectified by closely attending to how things come into words. A bottom-up focus on Bateson's sense of process may provide means for obtaining new and liberating experiential consequences from the Theory of Logical Types, instead of the oppressive consequences noted by Krippendorf [51].

4.1. Bateson on Convergent Stochastic Sequences

Bateson [53] identified the source of the problem when he recognized that " ... quantity and pattern are of different logical type and do not readily fit together in the same thinking." This is because "Not all numbers are the products of counting ... smaller ... commoner numbers ... are ... recognized as patterns in a single glance" [53]. Additionally, " ... we get numbers by counting or pattern recognition while we get quantities by measurement".

Bateson does not turn the problem around to ask how patterns in counting might explain the genesis of quantities, but he [53] does, however, devote considerable attention to the "generative process whereby the classes are created before they can be named". He (p. 216) follows Mittelstaedt in characterizing "a whole genus of methods of perfecting an adaptive act" in terms of feedback from a class of self-correcting practices, and in characterizing another genus of methods in terms of performance calibrations: " ... in these cases, 'calibration' is related to 'feedback' as higher logical type is related to lower."

Bateson [53] also identifies convergent "stochastic sequences" in the forms of feedback encountered in genetics and learning. This feedback is what makes prediction of future events possible. Although Bateson does not develop his thinking in this direction, he implies that these sequences should cohere in stochastic games of trial and error. These games inform a calibration language of measurement naming quantities embodied in metalinguistic instruments read in standardized units and predicted by metacommunicative theory. Devine [54] independently makes the same point, saying that:

> The computational mechanics approach [to stochastic processes] recognises that, where a noisy pattern in a string of characters is observed as a time or spatial sequence, there is the possibility that future members of the sequence can be predicted. The approach shows that one does not need to know the whole past to optimally predict the future of an observed stochastic sequence, but only what causal state the sequence comes from. All members of the set of pasts that generate the same stochastic future belong to the same causal state.

Rasch's [55,56] stochastic models for measurement are based on the same principle of reproducible patterns of invariance estimated from what are conceived as inherently incomplete samples of data collected in the past. These patterns are hypothesized to take various forms specified in a class of models as being structured by causal relationships retaining characteristic identities as repeatable constructs across samples of things, people,

or processes measured, and across samples of questions, indicators, or items comprising the measuring instruments [57–67].

Probabilistic models for measurement have been characterized in terms of complex adaptive systems [68], but are more commonly treated as scaling algorithms in the fields of psychometrics, educational measurement, survey research, etc. [69–77]. Recent collaborations of psychometricians and metrologists [65,78–86] are reaching beyond the usual analytic perspective to posit the viability of distributed interconnected ecosystem relations. These researchers suggest that systematic distinctions between numeric counts and measured quantities can be built into knowledge infrastructures in the same way that the standard kilogram unit for measuring mass has been globally adopted. Separating and balancing the utility of each logical type by providing the contextualization and maximizing their meaningfulness requires incorporating consensus standard methods for explicit experimental tests of, and theoretical justifications for, unit definitions and instrument calibrations.

4.2. Wright's Expansions on Rasch's Stochastic Models for Measurement

Echoing Bateson's language of the predictability obtained from convergent stochastic sequences, Wright [76] couches the advantages of Rasch's models in terms of their support for inferential processes of learning from the data in hand to predict future data not yet obtained. Moving beyond a merely analytic process of statistical inference, Wright [76] speaks to the need and the basis for stable units of measurement, as in the history of the economics of taxation and trade informed by instruments read in universally accessible common metrics.

Wright's position is supported by mathematical proofs of Rasch's [55] parameter separability theorem and decades of experimental evidence [87–89]. Rasch's models satisfy the axioms of conjoint additivity [90], and employ observed scores as minimally sufficient (i.e., both necessary and sufficient) estimators of the model parameters [91–94]. Statistical sufficiency is key here. Following Fisher [95,96], estimates that contain the complete information available in the data are necessary to models positing separable parameters [92]. Sufficiency is established mathematically as defining a set of invariant rules that is "an essentially complete subclass of the class of invariant rules" [97,98].

Insight into what this means can be gained by picking up where Bateson left off in his realizations that counting patterns are of a different logical type than quantities, and that feedback patterns aggregate to calibrate performances at a discontinuous new level of complexity. Wright [94,99–101] describes how patterns in counts can be experimentally evaluated relative to a probabilistic model requiring structural invariance and sufficient statistics to support or refute the hypothesis that a construct can be measured in an interval unit quantity. Where Bateson [53] stops with a distinction between being able to possess exactly three tomatoes but never being able to have exactly three gallons of water, Wright [76,77,94,99–101] shows how to derive what we want (measures and quantities) from what we have (counts and numbers).

Counts of events, such as correct answers to questions, or counts of categorical ratings, are typically interpreted as indicating the presence or absence of some social or psychological trait, an ability, knowledge, performance capacity, or behavior. Identical scores (summed counts or ratings) from the same items are interpreted as meaning the same thing, even if they may in fact conceal quite different patterns. Scores are known to be interpretable only in terms of an ordinal scale, one in which higher numbers mean more (typically), but with no associated capacity for determining how much more. Despite this caveat, ordinal scores are nonetheless routinely employed in statistical comparisons assuming interval units [73,101–103], contributing to the widespread promulgation of cross-level fallacies in knowledge infrastructures.

Wright [100] implicitly expands on Bateson's point concerning the difference between counts and quantities. He notes that oranges vary in size and juice content in such a way that it is not possible to tell how many oranges it will take to obtain two glasses (about a

pint, or almost 500 mL) of juice. "Sometimes it's only three. Other times it can take six." He finds that oranges are about half juice, by weight, and because "a pint's a pound the whole world round," a pint of orange juice will typically be produced by two pounds (about 1 kg) of oranges. His grocery store sells oranges in four-pound bags; therefore, he knows to take out half of the oranges in the bag to obtain a pint of orange juice.

The same principle applies to the difference between counts and quantities in any science. Figures 1 and 2 illustrate the practical difference between the kinds of comparisons supported by counts and measures. Figure 1 shows how the meanings of counts are restricted to the frame of reference defined by a particular group of persons responding to a particular group of questions. Imagine that a group of people all answer every question on a long test or survey, each item is scored as a count of correct or agreeable responses, and the items are then divided into two groups. One group is made up of all the highest scoring items, and the other is made up of all the lowest scoring questions. Each person then has two scores, one from the easy questions, and another from the hard ones.
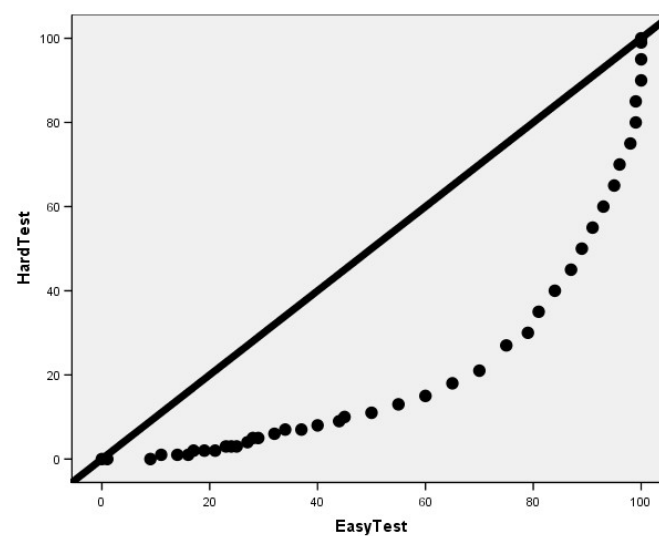
Figure 1. Percentage scores from numeric counts of responses to easy and hard questions.
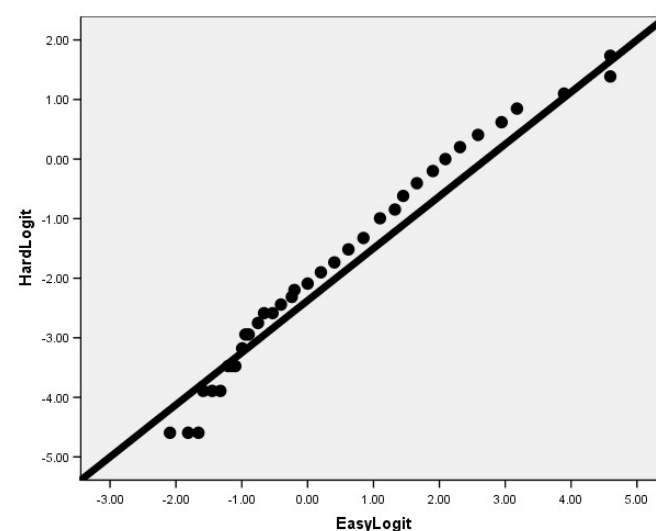
Figure 2. Log-odds unit (logit) measures from easy and hard questions.

Figure 1 is a plot of these pairs of scores, expressed as percentages of the maximum possible. Two points, at 0 and 100, are on the identity line, showing that the same score is obtained on both instruments. When someone obtains the highest possible score of 100 on

the hard questions, they also obtain the highest possible score on the easiest questions. Conversely, when someone obtains the lowest possible score of 0 on the easy questions, they also obtain a 0 on the hard ones. However, by definition, in between these extremes, everyone will have higher scores on the easy questions than on the hard ones.

In keeping with the ordinal status of the scores, the distance between adjacent scores changes depending on where they are positioned between 0 and 100. The relation between score changes, such that 1–1 and 99–99 indicate close correspondences, but the entire range of 25 to 100 on the hard questions spans only 80 to 100 on the easy questions. The same kind of relationship would be obtained if responses from a group of people were scored, divided into high and low groups, and pairs of scores were then assigned to all of the items and plotted.

Figure 2 shows what happens when something as simple as taking the log-odds of the response probabilities is used to estimate measures. Here, there is a nearly linear relationship between the pairs of measures from the two instruments, providing a basis for equating them to a common unit. More sophisticated estimation methods provide closer correspondences than this simple approximation [104]. Even so, Figure 2 illustrates why, as noted by Narens and Luce [105], Stevens was justified in adding log-interval scales as a fifth type in his measurement taxonomy (the others being nominal, ordinal, interval, and ratio).

Figure 1 shows the meaning of the same count to change nonlinearly from the easy questions to the hard ones. In addition, the unit size changes within each set of scores, so there is no way to identify a common score structure across them. Figure 2, however, shows that the two instruments could be scaled to a shared unit by estimating the item locations together via the common sample. Once they are equated and anchored at fixed values, the easy and hard questions, or any combination of subsets of them, can be administered to estimate measures in a stable, comparable quantitative metric [106,107].

Adaptive instrument administrations of this kind can be accomplished via paper and pencil or electronic self-scoring forms that report numeric patterns contextualized by measured quantities, along with uncertainty estimates. These forms provide immediate feedback at the point of use with no need for data computerization or analysis [108,109]. The universal availability of such instruments traceable to common metrics is essential to the creation of effective knowledge infrastructures separating forms of information by logical type.

Figure 3 shows the ogival relation between numeric scores and measured quantities, given data in a complete person-by-item matrix where everyone answers all questions. As can also be discerned in Figure 1, it is apparent in Figure 3 that, relative to their corresponding logits, scores at the extremes are packed more closely together than scores in the middle of the range. The same small percentage change on the vertical axis near 0 or 100 percent translates into a large logit difference relative to the smaller logit differences observed in the middle of the scale for the same percentage change.
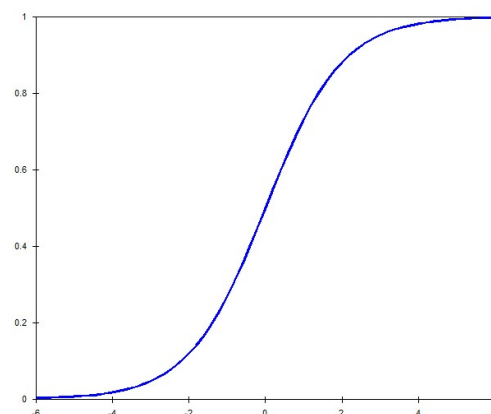
Figure 3. A logistic ogive illustrating the score–measure relationship.

Figures 4 and 5 show that linearly comparable results are obtained across samples of items and persons. Specifically, the structural invariance of the construct measured retains its characteristic properties in the manner required for the definition of meaningful quantity. Independent samples of items measuring the same construct will result in linear and highly correlated pairs of measurements, as shown in Figure 4. Conversely, independent samples of persons participating in the same construct will result in linear and highly correlated pairs of item calibrations, as shown in Figure 5.
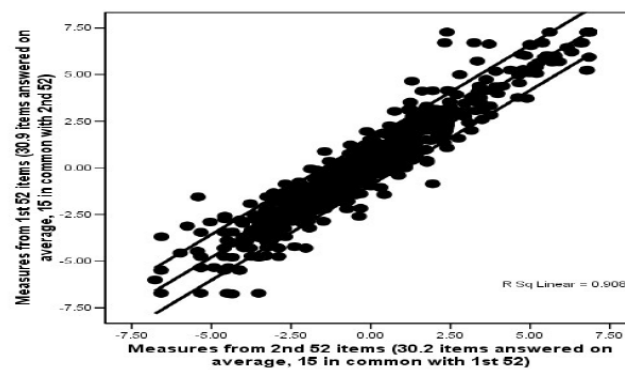
Figure 4. Statistically identical person measurements from separate but overlapping groups of items [110].
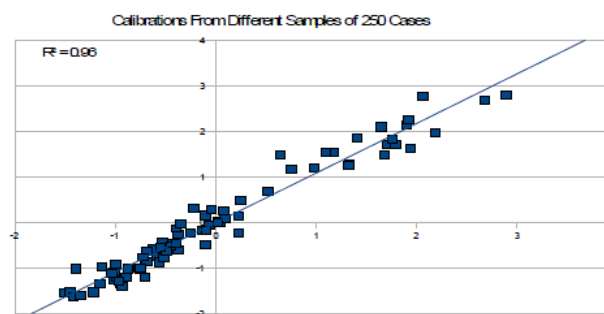
Figure 5. Statistically identical item calibrations from independent samples of respondents [110].

Finally, Figure 6 illustrates an example of the extent to which theoretical predictions implemented via an explanatory model of item components are borne out by empirical calibrations. Far from being rare exceptions to the rule in psychological and social measurement, as claimed by Sijtsma [111], these kinds of results and the principles informing them have been in wide circulation for decades [60–63,65–67,112–116]. These models and their results should be more routinely incorporated in a metrological context for standard research and practice.

4.3. The Difference That Makes a Difference as a Model Separating Logical Types

A key accommodation of the nonlinearity and scale-dependence of numeric counts became built into a wide range of measurement applications in a number of fields by Likert's [117] observation that the middle range of the logistic ogive in Figure 3 is approximately linear. This assertion of the near-equivalence of counts and measures for complete data in this range provided an inexpensive and pragmatic alternative to Thurstone's [118] more difficult methods that focused on calibrating instruments not affected in their measuring function by the particulars of who is measured. It also led to the systematic incorporation of unresolvable double binds in the messaging systems of virtually all major institutions. That is, the policies and practices of education, healthcare, government, and business all rely on nonlinear and ordinal numeric counts and percentages that are presented and treated as though they are linear, interval measured quantities.
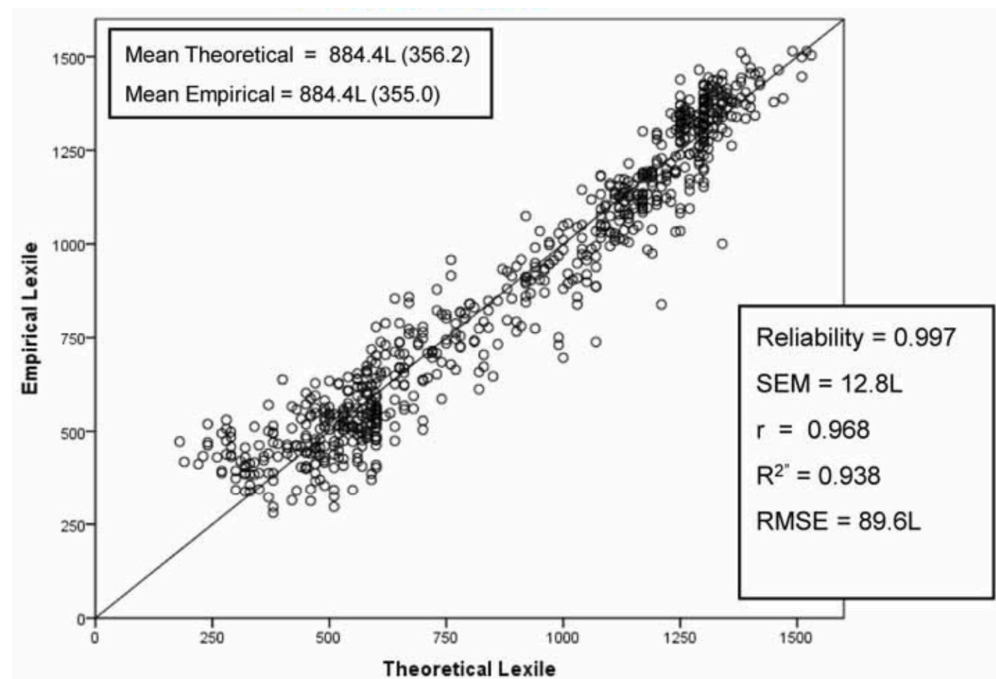
Figure 6. Empirical calibrations' correspondence with theoretical predictions [67].

What makes this a schizophrenic double bind? Just this: everyone involved is perfectly well aware that these counts of concrete events and entities (my ten small rocks or ten correct answers to easy questions, versus your five large rocks or five correct answers to difficult questions) are not quantitatively comparable, but everyone nonetheless plays along and acts as if they are. We restrict comparisons to responses to the same set of questions as one way of covering over our complicity in the maintenance of the illusion. This results in the uncontrolled proliferation of thousands of different, incomparable ways of purportedly measuring the same thing. Everyone accepts this as a necessity that must be accommodated—although it is nothing of the sort.

The schizophrenic dissonance of these so-called metrics is in no way incontrovertibly written in stone as an absolutely unrevisable linguistic formula. Following on from the emergence of key developments in measurement theory and practice in the 1960s, two highly reputable measurement theoreticians wrote in 1986 that (a) the same structure serving as a basis for measuring mass, length, duration, and charge also serves as a basis for measuring probabilities, and (b) that this broad scope of fundamental measurement theory's applicability was "widely accepted" [105]. However, if measurement approaches avoiding the epistemological error of confusing counts for quantities were "widely accepted" by 1986, why are communications infrastructures dependent on psychological and social measurements still so fragmented? Why have systematic implementations of the structures commonly found to hold across physical and psychosocial measurements not been devised so as to better inform the management of outcomes?

The answers to these questions likely involve the extreme difficulty of thinking, acting, and organizing in relation to multiple levels of complexity. As Wright concludes, the difference between counting right answers and constructing measures is the same as the difference between counting and weighing oranges. However, neither Wright nor Bateson ever mentions the infrastructural issues involved in making it possible to weigh oranges in a quality assured quantity value traceable to a global metrological standard. It is true that, just as one person with three oranges might have as much orange juice as someone else with six, so, too, might one person correctly answering three hard questions have more reading ability as someone else correctly answering six easier questions. However, this point alone contributes nothing to envisioning, planning, incentivizing, resourcing,

or staffing the development of the infrastructure needed to perform a measurement of reading ability, for instance, as universally interpretable as the measurement of mass.

An alternative examination of the counts–quantities situation is provided by Pendrill and Fisher [84], who apply engineering and psychometric models to numeric count estimates obtained from members of a South American indigenous culture lacking number words greater than three. The results clearly support Bateson's sense of the relation of counting to pattern recognition by demonstrating how intuitive visualizations approximate numeric tallies and so provide a basis for estimating quantitative measures. However, Pendrill and Fisher expand the consideration of the problem to a level of generality supporting the viability of a new class of metrological knowledge infrastructure standards.

The process begins by positing and testing the hypothesis that the only difference that makes a difference is the one between the ability of the person and the difficulty of the question. Of course, questions must be written and administered in ways that provide a reasonable basis for conducting these kinds of data tests [75,119]. Model fit may not be tenable when questions do not all tap the same construct, and when the specific question answered cannot be inferred from the response. Key to this instrument design process, therefore, is close attention to a theoretical rationale concerning what the construct is, how it varies from less to more, how the questions asked cause responses to manifest a particular invariance, and how the evidence accrued will warrant quantitative inferences.

When questions focus on specific cognitive or behavioral processes that can be inferred from the responses, and when the questions vary in their difficulty or agreeability across a wide range, response probabilities often consistently and monotonically increase from less to more. When enough questions are asked to reduce uncertainty to a small fraction of the observed variation [110,120], reproducible invariances can be reliably measured. Consistent reproduction of a unidimensional construct across persons and items results in the fit of data to the model, such that the log-odds of the probability P of a correct response to item i by person n depends only on the difference between the person measure B and the item calibration D:

$$\ln(P_{ni}/(1 - P_{ni})) = B_n - D_i \tag{1}$$

This model states a key difference that makes a difference in conceptually and operationally distinguishing numeric counts' level of logical type from that of measured quantities. The more able the person is, the more likely a correct response will be observed on any item, within the range of uncertainty. The more difficult an item is, the less likely a correct response will be observed, from any person, within the range of uncertainty.

Given model fit and an explanatory theory predicting item calibrations, ordinal numeric counts can be re-expressed as interval-measured quantities meaningfully interpreted in terms of differences that make a difference. Fit to the model provides empirical substantiation of the existence of unit quantities that stand as consistent, repeatable, and comparable differences. An explanatory theory successfully predicting the scale locations of items based on features systematically varied across them enables qualitative annotations to and interpretations of the quantitative scale [75,119].

Although this model unrealistically posits an unachievable ideal, it does so in the good company of the Pythagorean theorem, Newton's laws, and many other mathematical formulations of regularly reproducible patterns exhibiting the stability needed for predicting the future. Furthermore, the model, similarly to the Pythagorean, Newtonian, and many other examples from the history of science, also accommodates local idiosyncrasies in the context of an abstract standard explained by a predictive theory. The goal, therefore, is not to impose an artificially homogenized sameness but to instead contextualize individuals' creative improvisations to make them negotiable in the moment, in the manner of everyday language.

Extensive experience with models of these kinds in dozens of fields has resulted in sophisticated methods of parameter estimation, instrument equating, item banking, model fit analysis, reporting, etc. Given the patterns of human development over time, and of healing processes across a wide range of conditions, existing educational and healthcare

institutions are already designed around intuitions of the self-organizing invariant structures modeled and measured via Rasch's models. Seeing how one or more instruments designed to measure variation in a specific construct align item contents across samples is a convincing display of evidence as to construct stability [87,121–123]. These kinds of patterns led to a large study of reading comprehension in the 1970s involving 350,000 students in all 50 U.S. states, calibrating and equating seven major reading tests [124,125]. This work eventually led to the development of a theoretical model explaining item variation and the deployment of a reading comprehension metric [66,67,126] now reported annually for over 32 million students in the U.S. alone.

5. Discussion

In accord with Bateson, Wright [127] maintains that "As long as primitive counts and raw scores are routinely mistaken for measures . . . there is no hope of . . . a reliable or useful science." Wright and Linacre [101] point out that attention to the difference between counts and measures in psychological measurement is not new. Thorndike called for the relevant methods in 1904, and Thurstone devised approximate solutions in the 1920s. As was also shown by Luce and Tukey [128] for their new form of fundamental measurement, Rasch's solution, formulated as an extension of Maxwell's analysis of Newton's second law of motion, is both necessary and sufficient, no matter what science is involved [90–94,129–131].

This is as it should be if we aspire to knowledge infrastructures incorporating fair comparisons living up to the meaning of the balance scale as a symbol of justice. If observed counts summarize all of the available information for each person responding (or object rated, etc.), then those counts represent a pattern supporting the estimation of a quantity and associated uncertainty. Wright [77] says:

> Inverse probability reconceives our raw observations as a probable consequence of a relevant stochastic process with a useful formulation. The apparent determinism of formulae like F = MA depends on the prior construction of relatively precise measures of F and M . . . The first step from raw observation to inference is to identify the stochastic process by which an inverse probability can be defined. Bernoulli's binomial distribution is the simplest process. The compound Poisson is the stochastic parent of all such measuring distributions.

The history of science can be seen as multiple instances of situations in which repeated observations of data patterns invariantly structured as independent reproductions of the same phenomenon led to explanatory predictive models and the calibration of instruments measuring in standard units [89,131]. From this perspective, experimental evidence of patterns of structural invariance in numeric counts supports meaningful expression in an explanatory model comprising a conceptual determination useful for calibrating instruments measuring in interval units.

When the process is completed and a useful abstraction has become a habitually used labor-saving tool in the economy of language, rediscovering how meaningfulness was achieved and distributed becomes a difficult process. Science and technology studies, however, repeatedly arrive at conclusions focusing on problems of logical types, cross-level fallacies, and levels of meaning [6,42,47,132–136]. No efforts to date in this area, however, have capitalized on Bateson's logical type distinctions between counting and measurement in developing knowledge infrastructures that include quantitative social and psychological data.

Developments in this direction follow Finkelstein [137,138], who presented measurement and measurement systems in terms of symbolically represented information. Metrological psychometrics research is explicitly oriented toward clear separations of denotative numeric and metalinguistic quantitative logical types [65,78–86,139]. Quantity is furthermore defined in metacommunicative theoretical terms as requiring explanatory construct models [60–63,67,126].

Logical type distinctions of these kinds can be made across a potentially infinite array of levels. Transitions across levels of cognitive and moral reasoning have long been a focus of scaling research in developmental psychology [70,140–145]. Probabilistic models for measurement have been used to show how different methods of numerically scoring developmental levels can be equated to common metric quantities.

A logical type distinction between numeric counts and measured quantities should be made not only in the bottom-up calibration of instruments, but also in the top-down reporting of measurements [139]. Formative applications of measures in education and healthcare benefit from response-level data reports communicating feedback on unique individual patterns in the context of the expected learning progression, developmental sequence, or healing trajectory. No data ever fit a model perfectly, but the point is usefulness, not truth [55,146]. Given a well-studied construct and well-calibrated instrumentation measuring to a reasonably small degree of uncertainty, even markedly inconsistent response patterns may convey important information useful to teachers' instructional planning, or clinicians' treatment planning.

In education, therefore, individualized instruction should not be based only on quantitative measures and associated uncertainties indicating where the student is positioned along the learning progression. It should also take into account evidence of special weaknesses and strengths, as shown in the numeric pattern of responses scored correct and incorrect [147–150]. Notable directions for ongoing research in this area break out separate but interrelated investigations taking place at these levels:

- Denotative data reports for formative feedback on concrete responses mapping individualized growth;
- Metalinguistic Wright maps used in psychometric scaling evaluations and abstract instrument calibrations;
- Metacommunicative construct maps and specification equations illustrating explanatory models and formal predictive theories.

All three of these areas are supported by various software packages, such as the BEAR Assessment System Software [151–155], ConQuest [156], RUMM [157], Winsteps [158], various R statistics packages [159], and others.

6. Conclusions

Probabilistic models for measurement inform a multilevel context operationalizing a science of data, instruments, and theory implementing Bateson's emphasis on differences that make a difference. By hypothesizing a pragmatic ideal condition in which nothing influences a person's response to a question but the difference between their ability and the difficulty of the challenges posed, Rasch makes Bateson's maxim testable in a wide range of situations.

Wright laid out the means for understanding Rasch's models in ways conforming to Bateson's distinctions between the levels of complexity in meaning. Number is not quantity, and counting is not measurement; therefore, knowledge infrastructures of all kinds treating numeric counts as quantitative measures are compromised by avoidable cross-level fallacies and schizophrenic double binds. However, patterns in numeric counts can be—and are—used to test hypotheses of quantitative structure useful in constructing measures. Knowledge infrastructures respecting the limits of different logical types are not only possible, but have, in limited instances, already been created [67,78,123,126].

These infrastructures satisfy Whitehead's [160] recognition that "Civilization advances by extending the number of important operations which we can perform without thinking about them" (p. 61). Given the complexity of the problem, it is unreasonable to expect individuals to master the technical matters involved. This insight contradicts those, such as Kahneman [161], who recommend learning to think slower instead of allowing language's fast automatic associations to play out. Whitehead [160] held that this position, "that we should cultivate the habit of thinking of what we are doing," is "a profoundly erroneous truism." Developmental psychology [162–166] and social studies of science [167–169] have

arrived at much the same conclusion, noting that individual intelligence is often a function of the infrastructural scaffolding provided by linguistic and scientific standards. When these standards do not include the concepts and words needed for dealing with the concrete problems of the day, individuals and institutions must inevitably fail.

The cultural schizophrenia caused by communications routinely comprising mixed messages has led to dire and unsustainable consequences, globally. There are many who say immediate transformative efforts must be mounted in response to the urgency of the potentially catastrophic crises humanity faces. Few, if any, of those urging transformative programs speak to the most vital differences that matter in making a difference, those involving logical type distinctions sustained not at the level of individual skill and inclination, but at the institutional level of infrastructural standards.

This shift of focus from persuading and educating individuals to altering the environments they inhabit has been described as the difference between modernizing and ecologizing [167–169]. However, even when relational sociocognitive ecologies are conceived and examined [6,45,46,49,170,171], the value ascribed to metrology by Latour [172,173] is never connected with the corresponding need for new forms of computability addressing the design challenge of achieving results that are "both socially situated and abstract enough to travel across time and space" [6]. The focus on individuals is so deeply ingrained into the fabric of institutionalized assumptions and norms that it prevails even when shifting the environmental infrastructure is the ostensible object of inquiry [139].

Fears that repressive methods will be used to control populations, as has happened so often in the past, justify cautious evaluations of new proposals for innovative efforts to improve the human condition. Taking Scott's [174] suggestion of conceiving information infrastructures on the model of language converges with Hayek's [175] agreement with Whitehead [160] concerning the advancement of civilization, and with Hayek's associated suggestion that totalitarian impulses be countered by co-locating information and decision makers together in context, as noted by Hawken [176] (p. 21). This co-location opens up key opportunities for experimental trial and error approaches supporting noncoercive self-determination.

These opportunities are not easily obtained. The biological concept of ecosystem resilience, for instance, has been unenthusiastically received by social theorists because it appears to favor adaptive responses supporting the persistence of the status quo, and not the adaptive emergence of new forms of social life [177]. Similarly, ethnographic studies of culture often assume infrastructures function only to give material form to political realities that are uncritically taken to be permanent and stable [178]. Bottom-up approaches to conceptual speciation events, where new things come into words, may then provide an attractive alternative that informs infrastructuring as a kind of ontological experimentation [179].

Incorporating logical type distinctions in knowledge infrastructures may inspire more confidence in the capacity to balance tradition with creativity. In light of the connections between Bateson's and Wright's ideas on the differences between counting numbers and measured quantities, methodical sensitivity to cross-level fallacies and multiple levels of meaning may become an essential component of ontologically oriented infrastructure experiments deliberately intended to shape politics and power via participatory processes.

Perhaps clearer ways of formulating new classes of responses to these challenges will be found by recognizing and working with, instead of against, logical type discontinuities. As Oppenheimer [180] recognized, "all sciences arise as refinements, corrections, and adaptations of common sense." Similarly, Bohr [181] noted that mathematical models provide "a refinement of general language." Nersessian [182] shows how scientific research extends ordinary everyday cognitive practices. Gadamer [183] points out that the distinction between number and the noetic order of existence comprises "the world of ideas from which science is derived and which alone makes science possible" (pp. 35–36).

Bateson diagnosed the schizophrenic double-bind and cross-level fallacies built into knowledge infrastructures, emphasizing that counts are not measures. Wright expanded

on Rasch's mathematics to develop methods separating and balancing numeric counts and measured quantities, in the process extricating himself from what he perceived to be a schizophrenic situation. A psychometric metrology of emergent, bottom-up, self-organizing processes based in Bateson's, Rasch's, and Wright's works may provide the kind of experimental infrastructuring needed for generalized development of new kinds of outcomes for the wide array of institutions currently incorporating logical type confusions in their communications.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No original data are reported; source citations for all data-based figures are provided.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Russell, B. Mathematical logic as based on the theory of types. *Am. J. Math.* **1908**, *30*, 222–262. [CrossRef]
2. Whitehead, A.N.; Russell, B. *Principia Mathematica*; Cambridge University Press: Cambridge, UK, 1996.
3. Bateson, G. *Steps to an Ecology of Mind: Collected Essays in Anthropology, Psychiatry, Evolution, and Epistemology*; University of Chicago Press: Chicago, IL, USA, 1972.
4. Fisher, W.P., Jr. Wright, Benjamin D. In *SAGE Research Methods Foundations*; Atkinson, P., Delamont, S., Cernat, A., Sakshaug, J.W., Williams, R., Eds.; Sage Publications: Thousand Oaks, CA, USA, 2020.
5. Wilson, M.; Fisher, W.P., Jr. (Eds.) *Psychological and Social Measurement: The Career and Contributions of Benjamin D Wright*; Springer: Cham, Switzerland, 2017.
6. Star, S.L.; Ruhleder, K. Steps toward an ecology of infrastructure: Design and access for large information spaces. *Inf. Syst. Res.* **1996**, *7*, 111–134. [CrossRef]
7. Bowker, G.C. Susan Leigh Star Special Issue. *Mind Cult. Act.* **2015**, *22*, 89–91. [CrossRef]
8. Bowker, G.C. How knowledge infrastructures learn. In *Infrastructures and Social Complexity: A Companion*; Harvey, P., Jensen, C.B., Morita, A., Eds.; Routledge: New York, NY, USA, 2016; pp. 391–403.
9. Alker, H.R. A typology of ecological fallacies. In *Quantitative Ecological Analysis in the Social Sciences*; Dogan, M., Rokkan, S., Eds.; MIT Press: Cambridge, MA, USA, 1969; pp. 69–86.
10. Rousseau, D.M. Issues of level in organizational research: Multi-level and cross-level perspectives. *Res. Organ. Behav.* **1985**, *7*, 1–37.
11. Deleuze, G.; Guattari, F. *Anti-Oedipus: Capitalism and Schizophrenia*; Penguin: New York, NY, USA, 1977.
12. Brown, N.O. *Love's Body*; University of California Press: Berkeley, CA, USA, 1966.
13. Wright, B.D. On behalf of a personal approach to learning. *Elem. Sch. J.* **1958**, *58*, 365–375. [CrossRef]
14. Wright, B.D.; Bettelheim, B. Professional identity and personal rewards in teaching. *Elem. Sch. J.* **1957**, 297–307. [CrossRef]
15. Townes, C.H.; Merritt, F.R.; Wright, B.D. The pure rotational spectrum of ICL. *Phys. Rev.* **1948**, *73*, 1334–1337. [CrossRef]
16. Wright, B.D. *A Simple Method for Factor Analyzing Two-Way Data for Structure*; Social Research Inc.: Chicago, IL, USA, 1957.
17. Wright, B.D.; Evitts, S. Multiple regression in the explanation of social structure. *J. Soc. Psychol.* **1963**, *61*, 87–98. [CrossRef] [PubMed]
18. Wright, B.D.; Loomis, E.; Meyer, L. The semantic differential as a diagnostic instrument for distinguishing schizophrenic, retarded, and normal pre-school boys. *Am. Psychol.* **1962**, *17*, 297.
19. Wright, B.D.; Loomis, E.; Meyer, L. Observational Q-sort differences between schizophrenic, retarded, and normal preschool boys. *Child Dev.* **1963**, *34*, 169–185. [CrossRef]
20. Wright, B.D. Georg Rasch and measurement. *Rasch Meas. Trans.* **1988**, *2*, 25–32.
21. Bateson, G.; Jackson, D.D.; Haley, J.; Weakland, J. Toward a theory of schizophrenia. *Behav. Sci.* **1956**, *1*, 251–264. [CrossRef]
22. Boundas, C.V. (Ed.) *Schizoanalysis and Ecosophy: Reading Deleuze and Guattari*; Bloomsbury Publishing: London, UK, 2018.
23. Dodds, J. *Psychoanalysis and Ecology at the Edge of Chaos: Complexity Theory, Deleuze, Guattari and Psychoanalysis for a Climate in Crisis*; Routledge: New York, NY, USA, 2012.
24. Robey, D.; Boudreau, M.C. Accounting for the contradictory organizational consequences of information technology: Theoretical directions and methodological implications. *Inf. Syst. Res.* **1999**, *10*, 167–185. [CrossRef]
25. Cullin, J. Double bind: Much more than just a step 'Toward a Theory of Schizophrenia'. *Aust. N. Z. J. Fam. Ther.* **2013**, *27*, 135–142. [CrossRef]
26. Olson, D.H. Empirically unbinding the double bind: Review of research and conceptual reformulations. *Fam. Process* **1972**, *11*, 69–94. [CrossRef]

27. Radman, A. Double bind: On material ethics. In *Schizoanalysis and Ecosophy: Reading Deleuze and Guattari*; Boundas, C.V., Ed.; Bloomsbury Publishing: London, UK, 2018; pp. 241–256.

28. Stagoll, B. Gregory Bateson (1904–1980): A reappraisal. *Aust. N. Z. J. Psychiatr.* **2005**, *39*, 1036–1045.

29. Visser, M. Gregory Bateson on deutero-learning and double bind: A brief conceptual history. *J. Hist. Behav. Sci.* **2003**, *39*, 269–278. [CrossRef]

30. Watzlawick, P. A review of the double bind theory. *Fam. Process* **1963**, *2*, 132–153. [CrossRef]

31. Star, S.L. The structure of ill-structured solutions: Boundary objects and heterogeneous distributed problem solving. In *Boundary Objects and Beyond: Working with Leigh Star*; Bowker, G., Timmermans, S., Clarke, A.E., Balka, E., Eds.; The MIT Press: Cambridge, MA, USA, 2015; pp. 243–259.

32. Miller, J.G. *Living Systems*; McGraw Hill: New York, NY, USA, 1978.

33. Subramanian, S.V.; Jones, K.; Kaddour, A.; Krieger, N. Revisiting Robinson: The perils of individualistic and ecologic fallacy. *Int. J. Epidemiol.* **2009**, *38*, 342–360. [CrossRef]

34. Van de Vijver, F.J.; Van Hemert, D.A.; Poortinga, Y.H. (Eds.) *Multilevel Analysis of Individuals and Cultures*; Psychology Press: New York, NY, USA, 2015.

35. Palmgren, E. A constructive examination of a Russell-style ramified type theory. *Bull. Symb. Log.* **2018**, *24*, 90–106. [CrossRef]

36. Lee, S.; Chang, J.Y.; Moon, R.H. Conceptualization of emergent constructs in a multilevel approach to understanding individual creativity in organizations. In *Individual Creativity in the Workplace*; Reiter-Palmon, R., Kennel, V.L., Kaufman, J.C., Eds.; Academic Press: San Diego, CA, USA, 2018; pp. 83–100.

37. Maschler, Y. Metalanguaging and discourse markers in bilingual conversation. *Lang. Soc.* **1994**, *23*, 325–366. [CrossRef]

38. Kaiser, A. Learning from the future meets Bateson's levels of learning. *Learn. Org.* **2018**, *25*, 237–247. [CrossRef]

39. von Goldammer, E.; Paul, J. The logical categories of learning and communication: Reconsidered from a polycontextural point of view: Learning in machines and living systems. *Kybernetes* **2007**, *36*, 1000–1011. [CrossRef]

40. Volk, T.; Bloom, J.W.; Richards, J. Toward a science of metapatterns: Building upon Bateson's foundation. *Kybernetes* **2007**, *36*, 1070–1080. [CrossRef]

41. Blok, A.; Nakazora, M.; Winthereik, B.R. Infrastructuring environments. *Sci. Cult.* **2016**, *25*, 1–22. [CrossRef]

42. Crabu, S.; Magaudda, P. Bottom-up infrastructures: Aligning politics and technology in building a wireless community network. *Comput. Support. Coop. Work* **2018**, *27*, 149–176. [CrossRef]

43. Guribye, F. From artifacts to infrastructures in studies of learning practices. *Mind Cult. Act.* **2015**, *22*, 184–198. [CrossRef]

44. Hanseth, O.; Monteiro, E.; Hatling, M. Developing information infrastructure: The tension between standardization and flexibility. *Sci. Technol. Hum. Values* **1996**, *21*, 407–426. [CrossRef]

45. Hutchins, E. Cognitive ecology. *Top. Cogn. Sci.* **2010**, *2*, 705–715. [CrossRef]

46. Hutchins, E. Concepts in practice as sources of order. *Mind Cult. Act.* **2012**, *19*, 314–323. [CrossRef]

47. Karasti, H.; Millerand, F.; Hine, C.M.; Bowker, G.C. Knowledge infrastructures: Intro to Part, I. *Sci. Technol. Stud.* **2016**, *29*, 2–12.

48. Rolland, K.H.; Monteiro, E. Balancing the local and the global in infrastructural information systems. *Inf. Soc.* **2002**, *18*, 87–100. [CrossRef]

49. Shavit, A.; Silver, Y. "To infinity and beyond!": Inner tensions in global knowledge infrastructures lead to local and pro-active 'location' information. *Sci. Technol. Stud.* **2016**, *29*, 31–49. [CrossRef]

50. Vaast, E.; Walsham, G. Trans-situated learning: Supporting a network of practice with an information infrastructure. *Inf. Syst. Res.* **2009**, *20*, 547–564. [CrossRef]

51. Krippendorff, K. [Review] Angels Fear: Toward an Epistemology of the Sacred, by Gregory Bateson and Mary Catherine Bateson. New York: MacMillian, 1987. *J. Commun.* **1988**, *38*, 167–171.

52. Ladd, H.F. No Child Left Behind: A deeply flawed federal policy. *J. Policy Anal. Manag.* **2017**, *36*, 461–469. [CrossRef]

53. Bateson, G. *Mind and Nature: A Necessary Unity*; E. P. Dutton: New York, NY, USA, 1979.

54. Devine, S.D. *Algorithmic Information Theory: Review for Physicists and Natural Scientists*; Victoria Management School, Victoria University of Wellington: Wellington, New Zealand, 2014.

55. Rasch, G. *Probabilistic Models for Some Intelligence and Attainment Tests*; University of Chicago Press: Chicago, IL, USA, 1980.

56. Rasch, G. On general laws and the meaning of measurement in psychology. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability: Volume IV: Contributions to Biology and Problems of Medicine*; Neyman, J., Ed.; University of California Press: Berkeley, CA, USA, 1961; pp. 321–333.

57. Commons, M.L.; Goodheart, E.A.; Pekker, A.; Dawson-Tunik, T.L.; Adams, K.M. Using Rasch scaled stage scores to validate orders of hierarchical complexity of balance beam task sequences. *J. Appl. Meas.* **2008**, *9*, 182–199. [PubMed]

58. Dawson, T.L. New tools, new insights: Kohlberg's moral reasoning stages revisited. *Int. J. Behav. Dev.* **2002**, *26*, 154–166. [CrossRef]

59. Dawson, T.L. Assessing intellectual development: Three approaches, one sequence. *J. Adult Dev.* **2004**, *11*, 71–85. [CrossRef]

60. De Boeck, P.; Wilson, M. (Eds.) *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach*; Springer: New York, NY, USA, 2004.

61. Embretson, S.E. A general latent trait model for response processes. *Psychometrika* **1984**, *49*, 175–186. [CrossRef]

62. Embretson, S.E. *Measuring Psychological Constructs: Advances in Model.-Based Approaches*; American Psychological Association: Washington, DC, USA, 2010.

63. Fischer, G.H. The linear logistic test model as an instrument in educational research. *Acta Psychol.* **1973**, *37*, 359–374. [CrossRef]
64. Fischer, K.W. A theory of cognitive development: The control and construction of hierarchies of skills. *Psychol. Rev.* **1980**, *87*, 477–531. [CrossRef]
65. Melin, J.; Cano, S.; Pendrill, L. The role of entropy in construct specification equations (CSE) to improve the validity of memory tests. *Entropy* **2021**, *23*, 212. [CrossRef]
66. Stenner, A.J.; Smith, M., III. Testing construct theories. *Percept. Mot. Skills* **1982**, *55*, 415–426. [CrossRef]
67. Stenner, A.J.; Fisher, W.P., Jr.; Stone, M.H.; Burdick, D.S. Causal Rasch models. *Front. Psychol. Quant. Psychol. Meas.* **2013**, *4*, 1–14. [CrossRef]
68. Fisher, W.P., Jr. A practical approach to modeling complex adaptive flows in psychology and social science. *Procedia Comp. Sci.* **2017**, *114*, 165–174. [CrossRef]
69. Andrich, D. *Rasch Models for Measurement*; Sage University Paper Series on Quantitative Applications in the Social Sciences; Sage Publications: Beverly Hills, CA, USA, 1988; Volume 07-068.
70. Bond, T.; Fox, C. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*, 3rd ed.; Routledge: New York, NY, USA, 2015.
71. Engelhard, G., Jr. *Invariant Measurement: Using Rasch Models in the Social, Behavioral, and Health Sciences*; Routledge Academic: New York, NY, USA, 2012.
72. Fisher, W.P., Jr.; Wright, B.D. (Eds.) Applications of probabilistic conjoint measurement. *Int. J. Educ. Res.* **1994**, *21*, 557–664.
73. Hobart, J.C.; Cano, S.J.; Zajicek, J.P.; Thompson, A.J. Rating scales as outcome measures for clinical trials in neurology: Problems, solutions, and recommendations. *Lancet Neurol.* **2007**, *6*, 1094–1105. [CrossRef]
74. Kelley, P.R.; Schumacher, C.F. The Rasch model: Its use by the National Board of Medical Examiners. *Eval. Health Prof.* **1984**, *7*, 443–454. [CrossRef]
75. Wilson, M. *Constructing Measures: An Item Response Modeling Approach*; Lawrence Erlbaum Associates: Mahwah, NJ, USA, 2005.
76. Wright, B.D. A history of social science measurement. *Educ. Meas. Issues Pract.* **1997**, *16*, 33–45. [CrossRef]
77. Wright, B.D. Fundamental measurement for psychology. In *The New Rules of Measurement: What Every Educator and Psychologist Should Know*; Embretson, S.E., Hershberger, S.L., Eds.; Lawrence Erlbaum Associates: Hillsdale, NJ, USA, 1999; pp. 65–104.
78. Cano, S.; Pendrill, L.; Melin, J.; Fisher, W.P., Jr. Towards consensus measurement standards for patient-centered outcomes. *Measurement* **2019**, *141*, 62–69. [CrossRef]
79. Mari, L.; Maul, A.; Irribara, D.T.; Wilson, M. Quantities, quantification, and the necessary and sufficient conditions for measurement. *Measurement* **2016**, *100*, 115–121. [CrossRef]
80. Mari, L.; Wilson, M. An introduction to the Rasch measurement approach for metrologists. *Measurement* **2014**, *51*, 315–327. [CrossRef]
81. Mari, L.; Wilson, M.; Maul, A. *Measurement across the Sciences*; Springer: Cham, Switzerland, 2021.
82. Pendrill, L. Man as a measurement instrument [Special Feature]. *J. Meas. Sci.* **2014**, *9*, 22–33.
83. Pendrill, L. *Quality Assured Measurement: Unification Across Social and Physical Sciences*; Springer: Cham, Switzerland, 2019.
84. Pendrill, L.; Fisher, W.P., Jr. Counting and quantification: Comparing psychometric and metrological perspectives on visual perceptions of number. *Measurement* **2015**, *71*, 46–55. [CrossRef]
85. Wilson, M.; Fisher, W.P., Jr. Preface: 2016 IMEKO TC1-TC7-TC13 Joint Symposium: Metrology across the Sciences: Wishful Thinking? *J. Phys. Conf. Ser.* **2016**, *772*, 011001.
86. Wilson, M.; Fisher, W.P., Jr. Preface of special issue, Psychometric Metrology. *Measurement* **2019**, *145*, 190. [CrossRef]
87. Fisher, W.P., Jr. Invariance and traceability for measures of human, social, and natural capital: Theory and application. *Measurement* **2009**, *42*, 1278–1287. [CrossRef]
88. Fisher, W.P., Jr. What the world needs now: A bold plan for new standards [Third place, 2011 NIST/SES World Standards Day paper competition]. *Stand. Eng.* **2012**, *64*, 3–5.
89. Fisher, W.P., Jr. Measurements toward a future SI: On the longstanding existence of metrology-ready precision quantities in psychology and the social sciences. In *Sensors and Measurement Science International (SMSI) 2020 Proceedings*; Gerlach, G., Sommer, K.-D., Eds.; AMA Service GmbH: Wunstorf, Germany, 2020; pp. 38–39.
90. Newby, V.A.; Conner, G.R.; Grant, C.P.; Bunderson, C.V. The Rasch model and additive conjoint measurement. *J. Appl. Meas.* **2009**, *10*, 348–354.
91. Andersen, E.B. Sufficient statistics and latent trait models. *Psychometrika* **1977**, *42*, 69–81. [CrossRef]
92. Andrich, D. Sufficiency and conditional estimation of person parameters in the polytomous Rasch model. *Psychometrika* **2010**, *75*, 292–308. [CrossRef]
93. Fischer, G.H. On the existence and uniqueness of maximum-likelihood estimates in the Rasch model. *Psychometrika* **1981**, *46*, 59–77. [CrossRef]
94. Wright, B.D. Rasch model from counting right answers: Raw scores as sufficient statistics. *Rasch Meas. Trans.* **1989**, *3*, 62.
95. Fisher, R.A. On the mathematical foundations of theoretical statistics. *Philos. Trans. R. Soc. Lond. A* **1922**, *222*, 309–368.
96. Fisher, R.A. Two new properties of mathematical likelihood. *Proc. R. Soc. A* **1934**, *144*, 285–307.
97. Arnold, S.F. Sufficiency and invariance. *Stat. Probab. Lett.* **1985**, *3*, 275–279. [CrossRef]
98. Hall, W.J.; Wijsman, R.A.; Ghosh, J.K. The relationship between sufficiency and invariance with applications in sequential analysis. *Ann. Math. Stat.* **1965**, *36*, 575–614. [CrossRef]

99. Wright, B.D. Thinking with raw scores. *Rasch Meas. Trans.* **1993**, *7*, 299–300.

100. Wright, B.D. Measuring and counting. *Rasch Meas. Trans.* **1994**, *8*, 371.

101. Wright, B.D.; Linacre, J.M. Observations are always ordinal; measurements, however, must be interval. *Arch. Phys. Med. Rehabil.* **1989**, *70*, 857–867. [PubMed]

102. Michell, J. Measurement scales and statistics: A clash of paradigms. *Psychol. Bull.* **1986**, *100*, 398–407. [CrossRef]

103. Sijtsma, K. Playing with data–or how to discourage questionable research practices and stimulate researchers to do things right. *Psychometrika* **2016**, *81*, 1–15. [CrossRef]

104. Linacre, J.M. Estimation methods for Rasch measures. *J. Outcome Meas.* **1998**, *3*, 382–405.

105. Narens, L.; Luce, R.D. Measurement: The theory of numerical assignments. *Psychol. Bull.* **1986**, *99*, 166–180. [CrossRef]

106. Barney, M.; Fisher, W.P., Jr. Adaptive measurement and assessment. *Annu. Rev. Organ. Psychol. Organ. Behav.* **2016**, *3*, 469–490. [CrossRef]

107. Lunz, M.E.; Bergstrom, B.A.; Gershon, R.C. Computer adaptive testing. *Int. J. Educ. Res.* **1994**, *21*, 623–634. [CrossRef]

108. Kielhofner, G.; Dobria, L.; Forsyth, K.; Basu, S. The construction of keyforms for obtaining instantaneous measures from the Occupational Performance History Interview Ratings Scales. *Occup. Particip. Health* **2005**, *25*, 23–32. [CrossRef]

109. Linacre, J.M. Instantaneous measurement and diagnosis. *Phys. Med. Rehabil. State Art Rev.* **1997**, *11*, 315–324.

110. Fisher, W.P., Jr.; Elbaum, B.; Coulter, W.A. Reliability, precision, and measurement in the context of data from ability tests, surveys, and assessments. *J. Phys. Conf. Ser.* **2010**, *238*, 12036. [CrossRef]

111. Sijtsma, K. Introduction to the measurement of psychological attributes. *Measurement* **2011**, *42*, 1209–1219. [CrossRef]

112. Graßhoff, U.; Holling, H.; Schwabe, R. Optimal designs for linear logistic test models. In *MODa9—Advances in Model.-Oriented Design and Analysis: Contributions to Statistics*; Giovagnoli, A., Atkinson, A.C., Torsney, B., May, C., Eds.; Physica-Verlag HD: Heidelberg, Germany, 2010; pp. 97–104.

113. Green, K.E.; Smith, R.M. A comparison of two methods of decomposing item difficulties. *J. Educ. Stat.* **1987**, *12*, 369–381. [CrossRef]

114. Kubinger, K.D. Applications of the linear logistic test model in psychometric research. *Educ. Psychol. Meas.* **2009**, *69*, 232–244. [CrossRef]

115. Latimer, S.L. Using the Linear Logistic Test Model to investigate a discourse-based model of reading comprehension. *Rasch Model. Meas. Educ. Psychol. Res.* **1982**, *9*, 73–94.

116. Prien, B. How to predetermine the difficulty of items of examinations and standardized tests. *Stud. Educ. Eval.* **1989**, *15*, 309–317. [CrossRef]

117. Likert, R. A technique for the measurement of attitudes. *Arch. Psychol.* **1932**, *140*, 5–55.

118. Thurstone, L.L. Attitudes can be measured. *Am. J. Sociol.* **1928**, XXXIII, 529–544. In *The Measurement of Values*; Midway Reprint Series; University of Chicago Press: Chicago, IL, USA, 1959; pp. 215–233.

119. Fisher, W.P., Jr. Survey design recommendations. *Rasch Meas. Trans.* **2006**, *20*, 1072–1074.

120. Linacre, J.M. Rasch-based generalizability theory. *Rasch Meas. Trans.* **1993**, *7*, 283–284.

121. Fisher, W.P., Jr. Thurstone's missed opportunity. *Rasch Meas. Trans.* **1997**, *11*, 554.

122. Fisher, W.P., Jr. Metrology note. *Rasch Meas. Trans.* **1999**, *13*, 704.

123. He, W.; Li, S.; Kingsbury, G.G. A large-scale, long-term study of scale drift: The micro view and the macro view. *J. Phys. Conf. Ser.* **2016**, *772*, 12022. [CrossRef]

124. Jaeger, R.M. The national test equating study in reading (The Anchor Test Study). *Meas. Educ.* **1973**, *4*, 1–8.

125. Rentz, R.R.; Bashaw, W.L. The National Reference Scale for Reading: An application of the Rasch model. *J. Educ. Meas.* **1977**, *14*, 161–179. [CrossRef]

126. Fisher, W.P., Jr.; Stenner, A.J. Theory-based metrological traceability in education: A reading measurement network. *Measurement* **2016**, *92*, 489–496. [CrossRef] [PubMed]

127. Wright, B.D. Common sense for measurement. *Rasch Meas. Trans.* **1999**, *13*, 704–705.

128. Luce, R.D.; Tukey, J.W. Simultaneous conjoint measurement: A new kind of fundamental measurement. *J. Math. Psychol.* **1964**, *1*, 1–27. [CrossRef]

129. Fisher, W.P., Jr. *Rasch, Maxwell's Method of Analogy, and the Chicago Tradition*; University of Copenhagen School of Business, FUHU Conference Centre: Copenhagen, Denmark, 2010.

130. Fisher, W.P., Jr.; Stenner, A.J. On the potential for improved measurement in the human and social sciences. In *Pacific Rim Objective Measurement Symposium 2012 Conference Proceedings*; Zhang, Q., Yang, H., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 1–11.

131. Fisher, W.P., Jr. The standard model in the history of the natural sciences, econometrics, and the social sciences. *J. Phys. Conf. Ser.* **2010**, *238*, 12016. [CrossRef]

132. Berg, M.; Timmermans, S. Order and their others: On the constitution of universalities in medical work. *Configurations* **2000**, *8*, 31–61. [CrossRef]

133. Galison, P.; Stump, D.J. *The Disunity of Science: Boundaries, Contexts, and Power*; Stanford University Press: Stanford, CA, USA, 1996.

134. Robinson, M. Double-level languages and co-operative working. *AI Soc.* **1991**, *5*, 34–60. [CrossRef]

135. O'Connell, J. Metrology: The creation of universality by the circulation of particulars. *Soc. Stud. Sci.* **1993**, *23*, 129–173. [CrossRef]

136. Star, S.L.; Griesemer, J.R. Institutional ecology, 'translations', and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907–1939. *Soc. Stud. Sci.* **1989**, *19*, 387–420. [CrossRef]

137. Finkelstein, L. Representation by symbol systems as an extension of the concept of measurement. *Kybernetes* **1975**, *4*, 215–223. [CrossRef]

138. Finkelstein, L. Widely-defined measurement—An analysis of challenges. *Meas. Concern. Found. Concepts Meas. Spec. Issue Sect.* **2009**, *42*, 1270–1277. [CrossRef]

139. Fisher, W.P., Jr. Contextualizing sustainable development metric standards: Imagining new entrepreneurial possibilities. *Sustainability* **2020**, *12*, 9661. [CrossRef]

140. Bond, T.G. Piaget and measurement II: Empirical validation of the Piagetian model. *Arch. Psychol.* **1994**, *63*, 155–185.

141. Commons, M.L.; Miller, P.M.; Goodheart, E.A.; Danaher-Gilpin, D. *Hierarchical Complexity Scoring System (HCSS): How to Score Anything*; Dare Institute: Cambridge, MA, USA, 2005.

142. Dawson, T.L.; Fischer, K.W.; Stein, Z. Reconsidering qualitative and quantitative research approaches: A cognitive developmental perspective. *New Ideas Psychol.* **2006**, *24*, 229–239. [CrossRef]

143. Mueller, U.; Sokol, B.; Overton, W.F. Developmental sequences in class reasoning and propositional reasoning. *J. Exp. Child Psychol.* **1999**, *74*, 69–106. [CrossRef]

144. Overton, W.F. Developmental psychology: Philosophy, concepts, and methodology. *Handb. Child Psychol.* **1998**, *1*, 107–188.

145. Stein, Z.; Dawson, T.; Van Rossum, Z.; Rothaizer, J.; Hill, S. Virtuous cycles of learning: Using formative, embedded, and diagnostic developmental assessments in a large-scale leadership program. *J. Integral Theory Pract.* **2014**, *9*, 1–11.

146. Box, G.E.P. Some problems of statistics of everyday life. *J. Am. Stat. Assoc.* **1979**, *74*, 1–4. [CrossRef]

147. Wright, B.D.; Mead, R.; Ludlow, L. *KIDMAP: Person-by-Item Interaction Mapping*; Research Memorandum 29; University of Chicago, Statistical Laboratory, Department of Education: Chicago, IL, USA, 1980; p. 6.

148. Chien, T.W.; Chang, Y.; Wen, K.S.; Uen, Y.H. Using graphical representations to enhance the quality-of-care for colorectal cancer patients. *Eur. J. Cancer Care* **2018**, *27*, e12591. [CrossRef]

149. Black, P.; Wilson, M.; Yao, S. Road maps for learning: A guide to the navigation of learning progressions. *Meas. Interdiscip. Res. Perspect.* **2011**, *9*, 1–52. [CrossRef]

150. Fisher, W.P., Jr.; Oon, E.P.-T.; Benson, S. Rethinking the role of educational assessment in classroom communities: How can design thinking address the problems of coherence and complexity? *Educ. Des. Res.* **2021**, *5*, 1–33.

151. Irribarra, T.D.; Freund, R.; Fisher, W.P., Jr.; Wilson, M. Metrological traceability in education: A practical online system for measuring and managing middle school mathematics instruction. *J. Phys. Conf. Ser.* **2015**, *588*, 12042. [CrossRef]

152. Wilson, M.; Scalise, K. Assessment of learning in digital networks. In *Assessment and Teaching of 21st Century Skills: Methods and Approach*; Griffin, P., Care, E., Eds.; Springer: Dordrecht, The Netherlands, 2015; pp. 57–81.

153. Wilson, M.; Sloane, K. From principles to practice: An embedded assessment system. *Appl. Meas. Educ.* **2000**, *13*, 181–208. [CrossRef]

154. Fisher, W.P., Jr.; Wilson, M. Building a productive trading zone in educational assessment research and practice. *Pensam. Educ. Rev. Investig. Educ. Latinoam.* **2015**, *52*, 55–78. [CrossRef]

155. Fisher, W.P., Jr.; Wilson, M. An online platform for sociocognitive metrology: The BEAR Assessment System Software. *Meas. Sci. Technol.* **2020**, *31*, 34006. [CrossRef]

156. Adams, R.J.; Wu, M.L.; Wilson, M.R. *ACER ConQuest: Generalized Item Response Modelling Software, Version 4*; Australian Council for Educational Research: Camberwell, Australia, 2015.

157. Andrich, D.; Sheridan, B.; Luo, G. *RUMM 2030: Rasch Unidimensional Models for Measurement*; RUMM Laboratory Pty Ltd.: Perth, Australia, 2017.

158. Linacre, J.M. *A User's Guide to WINSTEPS Rasch-Model*; Computer Program, v. 4.8.1; Winsteps: Beaverton, OR, USA, 2021.

159. Mair, P. Rasch goes Open Source: Rasch models within the R environment for statistical computing. In *Horizons in Computer Science Research. Vol. 5. Horizons in Computer Science Research*; Clary, T.S., Ed.; Nova Science Publisher: Hauppauge, NY, USA, 2012.

160. Whitehead, A.N. *An Introduction to Mathematics*; Henry Holt and Co.: New York, NY, USA, 1911.

161. Kahneman, D. *Thinking Fast and Slow*; Farrar, Straus and Giroux: New York, NY, USA, 2011.

162. Barab, S.A.; Plucker, J.A. Smart people or smart contexts? Cognition, ability, and talent development in an age of situated approaches to knowing and learning. *Educ. Psychol.* **2002**, *37*, 165–182. [CrossRef]

163. Commons, M.L.; Goodheart, E.A. Cultural progress is the result of developmental level of support. *World Futur. J. New Paradig. Res.* **2008**, *64*, 406–415. [CrossRef]

164. Fischer, K.W.; Farrar, M.J. Generalizations about generalization: How a theory of skill development explains both generality and specificity. *Int. J. Psychol.* **1987**, *22*, 643–677. [CrossRef]

165. Sutton, J.; Harris, C.B.; Keil, P.G.; Barnier, A.J. The psychology of memory, extended cognition, and socially distributed remembering. *Phenomenol. Cogn. Sci.* **2010**, *9*, 521–560. [CrossRef]

166. Vygotsky, L.S. *Mind and Society: The Development of Higher Mental Processes*; Harvard University Press: Cambridge, MA, USA, 1978.

167. Bradley, K.; Hult, A.; Cars, G. From eco-modernizing to political ecologizing. In *Sustainable Stockholm: Exploring Urban. Sustainability in Europe's Greenest City*; Metzger, J., Olsson, A.R., Eds.; Routledge: London, UK, 2013; pp. 168–194.

168. Fisher, W.P., Jr.; Stenner, A.J. Ecologizing vs modernizing in measurement and metrology. *J. Phys. Conf. Ser.* **2018**, *1044*, 12025. [CrossRef]

169. Latour, B. To modernise or ecologise? That is the question. In *Remaking Reality: Nature at the Millennium*; Braun, B., Castree, N., Eds.; Routledge: London, UK, 1998; pp. 221–242.

170. Berque, A. An enquiry into the ontological and logical foundations of sustainability: Toward a conceptual integration of the interface 'Nature/Humanity'. *Glob. Sustain.* **2019**, *2*, 1–10. [CrossRef]

171. Clark, A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* **2013**, *36*, 181–204. [CrossRef]

172. Latour, B. *Science in Action: How to Follow Scientists and Engineers Through Society*; Harvard University Press: New York, NY, USA, 1987.

173. Latour, B. *Reassembling the Social: An. Introduction to Actor-Network-Theory*; Clarendon Lectures in Management Studies; Oxford University Press: Oxford, UK, 2005.

174. Scott, J.C. *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*; Yale University Press: New Haven, CT, USA, 1998.

175. Hayek, F.A. *Individualism and Economic Order*; University of Chicago Press: Chicago, IL, USA, 1948.

176. Hawken, P. *Blessed Unrest: How the Largest Movement in the World Came into Being and Why No One Saw It Coming*; Viking Penguin: New York, NY, USA, 2007.

177. Olsson, L.; Jerneck, A.; Thoren, H.; Persson, J.; O'Byrne, D. Why resilience is unappealing to social science: Theoretical and empirical investigations of the scientific use of resilience. *Sci. Adv.* **2015**, *1*, e1400217. [CrossRef] [PubMed]

178. Diener, P. Quantum adjustment, macroevolution, and the social field: Some comments on evolution and culture [and comments and reply]. *Curr. Anthropol.* **1980**, *21*, 423–443. [CrossRef]

179. Jensen, C.B.; Morita, A. Introduction: Infrastructures as ontological experiments. *Ethnos* **2017**, *82*, 615–626. [CrossRef]

180. Oppenheimer, R. Analogy in science. *Am. Psychol.* **1956**, *11*, 127–135. [CrossRef]

181. Bohr, N. *Essays 1958–1962 on Atomic Physics and Human Knowledge*; John Wiley & Sons: New York, NY, USA, 1963.

182. Nersessian, N.J. Maxwell and "The method of physical analogy": Model-based reasoning, generic abstraction, and conceptual change. In *Reading Natural Philosophy: Essays in the History and Philosophy of Science and Mathematics*; Malament, D., Ed.; Open Court: Lasalle, IL, USA, 2002; pp. 129–166.

183. Gadamer, H.-G.; Smith, P.C. *Dialogue and Dialectic: Eight Hermeneutical Studies on Plato*; Yale University Press: New Haven, CT, USA, 1980.