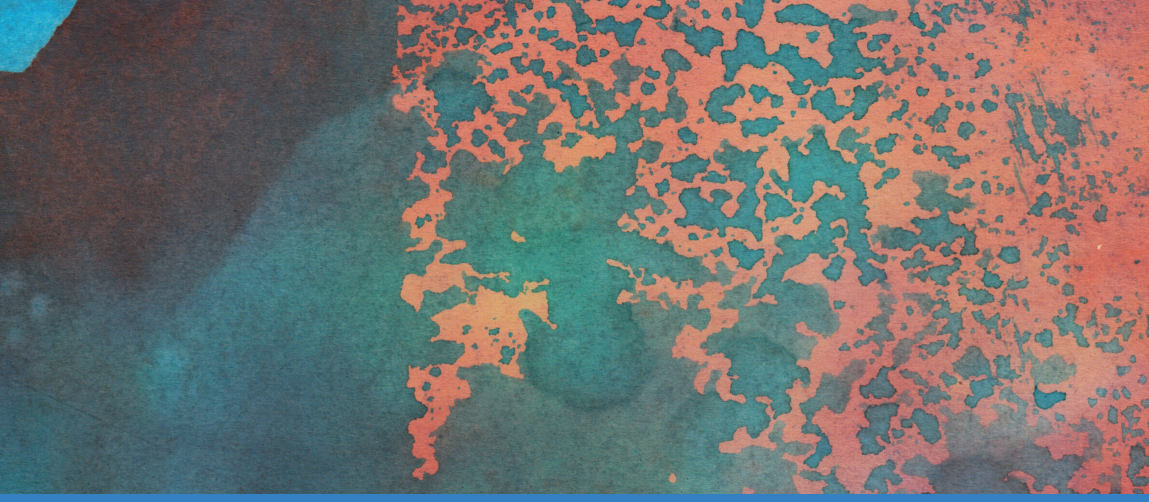
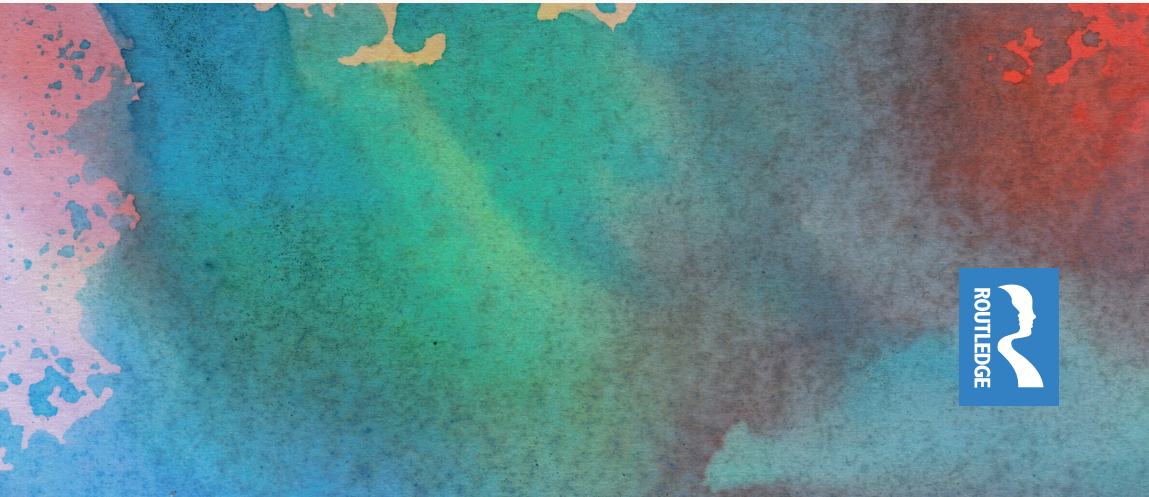


Routledge Research in Psychology

MIXED METHODS FOR PSYCHOLOGICAL MEASUREMENT

USING CRITICAL REALISM TO REFRAME INCOMMENSURABILITY

David F. Feldon



Mixed Methods for Psychological Measurement

This book presents a groundbreaking approach to measurement from a mixed methodological perspective, reframing the concept of incommensurability to harmonize qualitative and quantitative data in analyses.

It draws upon critical realism, latent variable theory, and phenomenography to illustrate how idiographic data can be incorporated into the development and validation of psychological measures. The text delves into foundational methodological assumptions, explores item and instrument validation techniques, and addresses issues of invariance. It further elucidates the application of idiographic strategies in conjunction with differential item function (DIF), item parameter drift (IPD), latent growth models, and the evaluation of measurement models. This innovative framework offers researchers robust tools for integrating diverse data types, enhancing the validity and reliability of their findings, and articulates ways in which these tools can be integrated into critical quantitative perspectives. It is an invaluable resource for anyone seeking to deepen their understanding of mixed methodologies and measurement.

It will appeal to scholars, researchers, and students whose goals are to integrate different modes of data, examine alternative perspectives on measurement, and apply new tools to psychological and social research.

David F. Feldon is a Professor of Instructional Technology and Learning Sciences at the Emma Eccles Jones College of Education and Human Services at Utah State University, USA.

Routledge Research in Psychology

This series offers an international forum for original and innovative research being conducted across the field of psychology. Titles in the series are empirically or theoretically informed and explore a range of dynamic and timely issues and emerging topics. The series is aimed at upper-level and postgraduate students, researchers, and research students, as well as academics and scholars.

Recent titles in the series include:

Unitary Developmental Theory and Psychological Development Across the Lifespan, Volume 1
A Model of Developmental Learning for Psychological Maturation and Recovery
Myles Sweeney

Unitary Developmental Theory and Organization Development, Volume 2
A Model of Developmental Learning for Change, Agility and Resilience
Myles Sweeney

Alan Watts in Late-Twentieth-Century Discourse
Commentary and Criticism from 1974-1994
Edited by Peter J. Columbus

COVID-19 and Psychological Distress in Africa
Communitarian Perspectives
Edited by Yamikani Ndasauka

For a complete list of titles in this series, please visit: https://www.routledge.com/Routledge-Research-in-Psychology/book-series/RRIP

Mixed Methods for Psychological Measurement

Using Critical Realism to Reframe Incommensurability

David F. Feldon

Routledge
Taylor & Francis Group
NEW YORK AND LONDON

First published 2024
by Routledge
605 Third Avenue, New York, NY 10158

and by Routledge
4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

ISBN: 978-0-367-70129-1 (hbk)
ISBN: 978-0-367-70132-1 (pbk)
ISBN: 978-1-003-14471-7 (ebk)

DOI: 10.4324/9781003144717

Typeset in Galliard
by KnowledgeWorks Global Ltd.

Contents

Preface

David F. Feldon

As academics, we often define who we are and what we do as much in opposition to other camps as on the specific merits of our own scholarly worldview. Accordingly, much of what we think we know is anchored in the reproduction of traditions and the ready consensus of other adherents to our original points of view—an intellectual echo chamber of formal and informal peer review. This book deliberately finds opportunities for convergence across paradigms to rethink foundational concepts. It builds an argument for a mixed methods approach to psychological measurement, bridging elements from philosophy, methodology, and statistics.

In doing so, it introduces several new approaches to psychometric analysis that center individual sensemaking and situativity. These approaches are not presented as final, authoritative models. Instead, they are proofs of concept for an alternative approach to measurement that inherently and explicitly rely on *qualia*—the "what it's like" of phenomena that differ across individuals—in addition to the distillation of natural and mathematical laws that attempt to describe phenomena in universal terms. From the stance of critical realism, these facets of psychological and social phenomena are equally real, and the perspective offered in this book is that measurement which engages and integrates both presents a more complete and valid view of the world than either one alone. In short, the goal of the book is to introduce a framework that differently conceptualizes both what can be measured and how it should be measured.

Contents of the Book

Chapter 1 engages the philosophical underpinnings of the work and establishes some basic terms and definitions. It addresses classic problems in the mixing of methods, such as the Quine–Duhem incompatibility thesis, which holds that quantitative and qualitative lenses for research cannot be reconciled within a single framework. The chapter reframes the problem, arguing that the purpose of research is not full and direct

representation of any phenomenon—rather, it is a disciplinary endeavor of data reduction to optimize the extraction of meaningful patterns from data to facilitate interpretation. Anchored in this perspective, the challenge of incommensurability becomes more tractable when we optimize data reduction strategies across methods to facilitate integration without losing either the richness of meaning often found in qualitative or the systematicity of quantitative measurement. Using the perspectives of critical realism and latent variable theory, the chapter suggests core strategies for capturing the causal relationships between phenomena that cannot be directly observed (i.e., latent constructs) and their indicators (i.e., manifest variables). These center on the ability to infuse the inclusion criteria for a given categorical variable with as much detail as needed to capture the nuance of traditional modes of qualitative inquiry, which creates a basis for establishing commensurability between representations of qualia and quantitative data.

Chapter 2 delves into the various ways in which personal sensemaking and individual experience shape both the development of scale items and the ways in which respondents formulate the selection of response options, infusing the process of measurement with constructed meaning that is not always considered. Conceptualizing such influences as latent constructs that do not hold inherently quantitative or qualitative properties, the chapter argues against the position of some measurement theorists that a given phenomenon may not be meaningfully measured if its asserted *a priori* nature is not quantitative. Expanding on this perspective, the chapter further invokes assumptions of phenomenography that the range of expressable human experience—while potentially vastly large—is not infinite, which permits sampling strategies to elicit meaningful information about a population. Having a rational and definable proportion of category frequency (i.e., not dividing by infinity) permits inferences about natural populations of qualitative phenomena necessary for measurement.

Chapter 3, authored with James Peugh, examines issues of stability, reliability, validity, and generalizability that arise when using qualitative data in measurement. Because categorical variables do not naturally have symmetrical and normally distributed measurement error, it has been difficult to argue that categorical variables could be appropriately integrated into models of continuous measurement. However, this chapter integrates several developments in the application of central limit theorem that theoretically permit the estimation of normalized measurement error for categorical variables by parsing sampling error from classification (i.e., rater) error. Based on this approach, we argue for the ways in which measures based on categorical data can be generalizable in multiple ways, integrating the generalizability logics of both quantitative and qualitative traditions. Specifically, we highlight the ways in which generalizability arguments typically used in relation

to the validity of measures align clearly with qualitative conceptions of analytic generalization and case-to-case transfer.

Chapter 4, authored with Kaylee Litson, applies the logic of qualitative member checking to interpreting residual variance in measurement models. Introducing a multiple-indicator multiple cause (MIMIC) model based on respondent estimates of alignment between their own sensemaking and their selection of fixed response options on a multiple choice assessment, we analyze and interpret structured residuals in the model to draw inferences about the ways in which personal sensemaking may influence the validity of specific assessment items on the basis of incommensurability between respondent knowledge and the response options available. One application of this approach is an evaluation of item-level validity to differentiate the extent to which misalignment between personal conceptions and constructed item response options can parse variance attributable to differences in knowledge and independent differences in item interpretation.

Chapter 5, authored with Kaylee Litson, Deborah Fields, Heather Clark, Lorraine Gale, Rebecca Brockbank, and Brinleigh Cahoon, demonstrates that individual sensemaking linked to experience and context can behave lawfully and be modeled over time using latent growth modeling. Leveraging categorical variables based on idiographic analysis of 548 interview transcripts, we examine the stability of residuals over time to assess the extent to which such variables behave lawfully both within individuals and in relation to trends in group variance. In doing so, we argue that the logic of nomological networks as a tool for evaluating validity can be applied to categorical variables derived from qualitative data. Further, we examined the convergence of meaning between the findings of the statistical modeling and qualitative case studies of individuals across the full distribution of the sample. Based on both the interindividual stability of semantic trajectories and the ability to map individual quantitative representations back to qualitatively distilled meanings, we conclude that this approach to modeling avoids concerns of incommensurability when representing qualitative phenomena using quantitative modeling.

Chapter 6, authored with Young Min Kim, Kaylee Litson, Noha Ramadan, Deborah Fields, and Rebecca Brockbank, integrates an idiographic categorical variable into a unidimensional measurement model as a moderator of both factor loadings and intercepts using moderated nonlinear factor analysis. Starting with an existing scale of research self-efficacy that exhibits item parameter drift (IPD; i.e., differential item function across repeated measurement instances), the incorporation of a idiographically derived categorical moderator theoretically linked to research self-efficacy mitigates IPD for multiple items. We argue that the categorical moderator should be considered an integral part of the measurement model for its ability to represent salient facets of individual sensemaking and social context as a

foundational part of the phenomenon that is the target of measurement. Shifting to a measurement model approach that incorporates qualia permits a more holistic representation of the underlying phenomenon, resulting in more usable measurement on the basis of stronger measurement invariance and more valid measurement based on a fuller representation of the target phenomenon.

Chapter 7 concludes this book by reframing the arguments presented in previous chapters to understand the importance of individual positionality and sensemaking in the context of critical quantitative perspectives. Criticisms of traditional approaches to measurement and quantitative analysis point out that the worldviews of measure-makers and the traditional tools of validation are likely to erase or corrupt meaning that does not align with the preconceived biases of those worldviews. In this circumstance, bias—including presumptions of white supremacy—prevent accurate and just representations of data. These biased representations can then be weaponized to protect the inequitable distribution of power and resources within society. Discussing the approaches to measurement models presented in Chapters 4–6 as tools to more accurately, equitably, and completely distill phenomena of interest, the chapter argues that these approaches may help advance the goals of critical perspectives using a quantitative lens, because they center individual sensemaking that is grounded in personal experiences and positionality.

Code for all analyses is provided in the appendix.

Future Directions

As noted above, the models and arguments presented here establish a foundation for a different approach to thinking about and enacting measurement of psychological phenomena. However, this foundation is not fully developed. It entails premises that provide logical consistency and permit a mixed methods approach to measurement to be entertained. However, the needs for extensive testing through application and simulation studies and the further development of new concepts remain. It is my hope that others will find the ideas presented worthy of such efforts in pursuit of better and more equitable tools for measurement.

Acknowledgements

As should be evident through the inclusion of many co-authors in a "sole authored" book, this work would not have been possible without the expansive support and tolerance of many people. Drs. Kaylee Litson and James Peugh have been remarkable colleagues whose vast statistical expertise has supported many of my own flights of fancy. Their deep knowledge and flexible thinking have advanced the notions presented in this book from utterances of "I've been thinking about…" to operationalized models and mathematical formulas that enable others to make sense of and engage these approaches. On the qualitative side of these endeavors, I owe a huge debt of gratitude to Dr. Deborah Fields who recognized the potential in the project and was willing to contribute her qualitative expertise to the analysis of a vast corpus of interview data, leading a dedicated team of junior scholars including Dr. Heather Clark, Rebecca Brockbank, Lorraine Jessop, Noha Ramadan, and Brinleigh Cahoon to support this work. Likewise, I am grateful to Drs. Robert Mislevy, Mario Suarez, and Annie Wofford who reviewed early versions of several chapters and were willing to provide their feedback and encouragement.

I also gratefully acknowledge the support of the National Science Foundation, whose support of my research on doctoral education has yielded the vast corpus of both qualitative and quantitative data used in Chapters 5 and 6. The full corpus of deidentified data are available in open source format on the Open Science Framework site at https://osf.io/hymus/. This material is based upon work supported under Awards 1431234, 1431290, and 1760894. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. I also thank my collaborators on these grants whose work helped conceptualize, gather, and make sense of these data: Drs. Michelle Maher, Kimberly Griffin, Josipa Roksa, Jennifer Blaney, and Annie Wofford.

1 Nomothetic and Idiographic

David F. Feldon

Most people who are passingly familiar with mixed methods are acquainted with the incommensurability hypothesis as discussed and rejected by Tashakkori and Teddlie (2003; Teddlie & Tashakkori, 2010) and Creswell (2003). Derived from the Quine–Duhem thesis and further elaborated in Kuhn's (1970) treatise on the nature of scientific argument, the gist is that diverging theories or paradigms cannot be reconciled, because the data collected and the meanings represented by their terminology are situated within distinct sets of assumptions that shape their development, application, and interpretation. Therefore, there can be no shared foundation from which to reconcile or synthesize conflicting viewpoints. Feyerabend's (1975) expanded articulation of incommensurability explicitly includes the incommensurability of divergent worldviews. To use Plato's metaphor, different theories, paradigms, or worldviews each "carve nature at its joints" but recognize and define those joints differently. In this regard, traditionally qualitative lenses (typically inductive, constructivist, and context-heavy) and traditionally quantitative lenses (typically deductive, realist, and context-light) cannot be meaningfully integrated, because they rest on divergent ontological and epistemological belief systems (Guba & Lincoln, 1981; Lincoln, 2010).

While many mixed methodologists reject the incommensurability hypothesis as irrelevant or unproductive—often marshalling Dewey's philosophical pragmatism in doing so—others have argued emphatically that setting aside these concerns without resolving them fails to complete necessary philosophical work in establishing a foundation for mixed methods. In the best case, the pragmatist rejection of a dualistic perspective permits "a discussion no longer crippled by unhelpful epistemological dichotomies … [but] unable to provide *the* philosophical foundation for mixed methods research" (Biesta, 2010, p. 114; emphasis in original). In the worst case, it fails to "make the premises, assumptions, and paradigmatic bases of our work clear, or worse yet, pretend[s] we have no premises, assumptions, or paradigmatic bases" (Lincoln, 2010, p. 5; see also Creamer, 2018). The

DOI: 10.4324/9781003144717-1

emerging perspective of dialectical pluralism argues that effective methodological mixing requires researchers to, on a case-by-case basis, "dialectically listen and consider multiple methodological concepts, issues, inquiry logics, and particular research methods and construct the appropriate mix for each research study" (Johnson, 2017, p. 167). However, the reconceptualization of research logics for every study fails to provide shared "premises, assumptions, or paradigmatic bases" (Lincoln) that can offer a foundation for conclusions to be directly compared across studies. Here, I offer a new perspective on strategies to make sense of mixed data within the same analysis by delving deeper into the nature of incommensurability as a metaphor and exploring its implications for measurement.

Framing Characteristics of Data and Methodologies

Qualitative and *quantitative* are most precisely descriptors of data (verbal vs. numeric), rather than descriptors of worldviews or methodological frameworks. It is not always the case that verbal data are used exclusively for inductive, situated analyses without an intent to generalize across a larger population. Likewise, numeric data can be used to characterize individuals in ways that are both inductive and not intended to characterize a larger group beyond the sample or individual participant. Accordingly, further discussion will utilize terms intended to more fully reflect the different assumptions characteristic of divergent paradigms: *idiographic* (emphasizing attention to individual cases without a driving value of generalizability) and *nomothetic* (emphasizing attention to lawful [in the sense of laws of nature] and generalizable commonalities across cases).

Although *idiographic* and *nomothetic* have both been used within the quantitative psychology community to refer to quantitative characterizations at the case and population levels (e.g., Molenaar, 2004; Nesselroade, Gerstorf, Hardy, & Ram, 2007), the philosophical origins of these terms are not based on the use of types of analytic tools or notions regarding meaningful sample size (Robinson, 2011). Windelband's (1894, 1901) introduction of these terms specified that idiographic knowledge describes and explains individual phenomena and nomothetic knowledge provides generalities that hold across a class of individual phenomena in the form of theories or laws. Accordingly, the case approach of idiographic analysis can and has historically included qualitative data in its analysis to more effectively conceptualize the phenomenological and narrative accounts of individuals (Runyan, 1983).

Whereas idiographic research can entail a focus on distilling the experiential characteristics of a phenomenon typically captured through qualitative data (e.g., transcripts of interviews, observations), there are numerous and widely varied nomothetic research strategies that likewise analyze verbal

or observational data without valuing personal sensemaking or perspective (e.g., *quantitizing* to facilitate statistical analyses). Conversely, numeric data are traditionally the focal mode for nomothetic research, but numbers can be imbued with highly situated meanings that contribute directly to individuals' sensemaking around experienced phenomena (e.g., *qualitizing* to imbue a trend, latent class, or factor with a semantically descriptive meaning). Because interpretive frames require anchors to specific ontological and epistemological stances that simple data modalities do not, the traditionally conceived incompatibility between methodological approaches is most effectively framed around the idiographic or nomothetic objectives of inquiry.

Incommensurability

The concept of incommensurability originates from mathematics as a property of a relation between two entities for which their ratio cannot yield a rational number. In short, this does not mean that the core entities (quantities or line segment lengths) cannot be productively or appropriately used together to provide meaningful information. Instead, when we attempt to put them into a unitary metric, there is information that we lose through approximation. For example, we know that if we take the diameter of a circle (a line segment with a definite length) and try to map it onto the circumference of the same circle, the length of the segment is equal to the circumference divided by π. The catch is that π is not a rational number, so we need to approximate it in order to use it. We round to $\pi \approx 22/7$ or $\pi \approx 3.14$, but both of those values leave out some greater or lesser amount of information. The infinite string of non-repeating decimals provides ever more precise estimates of the relation between the diameter and the circumference, but it is never complete—no matter how much analysis goes into it, there is always just a little bit more meaning that cannot be mapped between the two.

In nomothetic research, it is commonly understood that no measure can perfectly capture a phenomenon and that one source of measurement error comes specifically from limitations on its precision. For example, although Guttman (1944) argued that a score could not be part of a scale unless the data which gave rise to the score were represented in their totality by the score (i.e., deterministic conjoint additivity), Rasch's (1960) objectivity frames scores as stochastic. That is, scores can reflect probabilistic characteristics of the phenomenon within a true scale (Wright & Bell, 1984). In classical test theory, measurement error is defined as the extent to which an observed value deviates from a theoretical (and not directly obtainable) true score (Lord & Novick, 1968). From a more pragmatic perspective, the basic implementation of measurement entails rounding to the nearest useful or detectable unit as a practical necessity, for which measurement uncertainty is computed.

A comparable challenge has been identified within idiographic research as phenomenological researchers attempt to understand the meanings constructed by their research participants. If, within that tradition, understandings of reality—or reality itself—are constructed by each individual, then the precise meanings that they ascribe to words and events must be to some extent unique and distinct from those constructed by the person trying to understand their experiences. Drawing on Husserl's work on the problem of interpersonal understanding, idiographic researchers may engage in iterative dialogue with an individual to construct a meaningful representation of the other's perspective by successively refining their own understanding of what the individual means by specific utterances. The assumption is that perfect alignment will never be attained, but that intersubjectivity can be sufficiently close to permit an appreciation of the individual's perspective and lived experience in their knowledge construction (Zahavi, 2007).

From a poststructural perspective (Derrida, 1967/1976; Roth, 2007), meaning derived from others' use of language as an expression of their intended meaning is, likewise, not knowable. Derrida argued that the selection of articulated words inherently precludes the simultaneous use of alternative words that would add additional information about the speaker's intended meaning. Further, the broader linguistic context for an utterance cannot be fully reconstructed. Thus, communication constrains the ways in which intended meaning might be expressed in the moment and understood after the fact. Accordingly, a given moment of communication (i.e., qualitative data) can be understood by anyone other than the speaker themselves only as a social and relational experience that is incommensurate with the utterance itself or even the retrospectively articulated intention of the speaker.

In this way, the interpretation of incommensurability in its mathematical sense presents a well-established and addressable challenge within both idiographic and nomothetic traditions. How does this perspective help to resolve the dilemma that the more conventional incommensurability hypothesis poses for mixed methodological work? I argue that it provides a conceptual common ground from which methodological mixing might proceed.

Indeed, Kuhn (1976, p. 191) states explicitly that incommensurability does not entail incompatibility:

> Most readers have supposed that when I spoke of theories as incommensurable, I meant that they could not be compared. But "incommensurability" is a term borrowed from mathematics, and it therefore has no such implication. The hypotenuse of an isosceles right triangle is incommensurable with its side, but the two can be compared to any required degree of precision. What is lacking is not comparability but a unit of length in terms of which both can be measured directly and exactly.

Research as Data Reduction

Fundamentally, all social science research is a process of data reduction. Using different methods and assumptions, both idiographic and nomothetic researchers collect data, distill those data into aggregate trends, and then interpret the manner and extent to which the attained abstractions add value to theoretical and practical understandings of problems of interest. Although different paradigms express different priorities for the extent to which data should be reduced for appropriate interpretation, it is worth noting that idiographic research is never simply a compilation of intact transcripts, and nomothetic research is never simply a set of measurement data. Further, neither idiographic nor nomothetic researchers would argue that their apprehension of the underlying phenomena of interest were perfectly captured in the data—some information from the world is always left untapped by even the most thorough empirical investigation.

Each mode of research engages a symbol system to facilitate the analysis of the data collected in service to a research question. In the construction and use of any symbol system for research, the question of how to handle incommensurability always arises: What is the most appropriate and effective way to encode and interpret data such that the amount of information lost through truncation is not fundamentally threatening to the ability to meaningfully answer the research question?

By engaging various strategies to conceptualize the extent to which meaningful information is truncated in the course of any mode of research, this book explores how the Quine–Duhem/Kuhn problem of incommensurability between the idiographic and the nomothetic might be resolved in the process of measurement. In developing strategies to gauge or manage the scope of incommensurability entailed in data collection and analysis, we can construct a framework that reduces the "macro" incommensurability challenges (i.e., worldviews) facing integration across idiographic and nomothetic paradigms by finding common structures to characterize the "micro" incommensurabilities that result from data collection within either paradigm.

Incommensurability is not merely a quantity of misspecification or uncertainty—it is a framing of the collective differences between a latent phenomenon and its manifest observations, regardless of the mode of inquiry. These differences are both structural and perspectival, such that incommensurability is neither specifically nomothetic nor idiographic; the quantitative and qualitative strategies used are merely different tools for accessing an underlying phenomenon, with latent properties not constrained in nature to the modes of observation used or their epistemological assumptions.

Without a clear and unified strategy for gauging the extent of incommensurability that may exist for given sets of nomothetic or idiographic data, the utility of this perspective may appear limited. However, Stevens' (1946) definition of categorical variables provides a valuable tool to sidestep this

impasse. Nominal values are based solely on our ability to separate data into qualitatively distinct groups and then assign those group identities arbitrary numeric values (provided that the category criteria are constructed such that any given datum can be appropriately assigned to one and only one group). Indeed, Michell (1999; 2008) critically points out that most psychological measurement merely reflects extrapolations on the frequency distributions of categories. Thus, whatever idiographic criteria we use to group data into codes, themes, or other groupings of thick description have a certain level of incommensurability with the underlying phenomena from which they were derived. In turn, this incommensurability is inherently conveyed to the characteristics of the corresponding nominal categorical variable.

From a nomothetic tradition, the relationships between quantities can be modeled mathematically, such that they represent meaningful relationships between the quantities and those constructs they represent. For example, a χ^2 analysis or Fisher's exact test can examine the probability that a specific joint frequency distribution differs significantly from that predicted by a specified joint frequency or by chance. Accordingly, such analyses can inform our understanding of lawful relationships between categorical variables without consideration of, or changes to, the underlying richness or complexity of the idiographic constructions that defined them. The characteristics of these categories include the nature and scope of incommensurability, which are conveyed to the nominal scale by the same analyses that gave rise to each category's membership criteria. With incommensurability necessarily identical between the idiographic and the quantified categorical variable by definition, it is no longer necessary to estimate the extent of incommensurability within each type of data for the purposes of trying to align them.[1]

Dichotomous variables represent a special case of nominal scales, because the relationship between binary values (i.e., 1=true, 0=false for the application of a specific set of category criteria to a given case) and other constructs can be assessed in ways that inherently yield only linear relationships with other constructs (Aldrich & Nelson, 1984). Thus, the predefined incommensurability inherited by the dichotomous variable from its ideographically based criteria can be mapped with a quantifiable precision onto other (quantitative) variables through statistical analyses that estimate an error or residual term. In other words, the idiographic category is fully commensurate with the linear combination of nomothetic constructs (including an error/residual term) in the statistical analysis.

Mixing Methodologies in Measurement

By definition, measurement is a quantitative endeavor typically aligned with nomothetic inquiry. The assignment of numeric values to specified instances of a target phenomenon requires a monotonic correspondence between

some property and a specified quantity, which can be interpreted identically across distinct instances and contexts (i.e., it is unidimensional and reliable). Notwithstanding several competing expansions of this idea within the realm of measurement theory (see Borsboom, 2005; Michell, 1999), there exists an apparent incompatibility between this core concept and worldviews associated with idiographic inquiry: For idiographic research, reality is constructed and interpreted through the active meaning-making of individuals, based on their own experiences and relationships to others. As a result, the notion that it is possible for a given score to be equivalently meaningful across people and situations is fundamentally problematic.

This concern is not exclusive to non-quantitative idiographic perspectives. Wright and Stone (1999, p. 34) observe that "in practice, unidimensionality is conceptual rather than factual, qualitative rather than quantitative, an idea and intuition rather than an experience. No actual test can be perfectly unidimensional." Further, Nesselroade and colleagues (2007; see also Kagan, 1980) challenged the fundamental assumption that identical responses to survey items by the same person at different points in time necessarily have the same meaning. Molenaar (2004) likewise points out that—even when all data are quantitative and meet traditional standards of measurement quality—the dynamics of development measured over time within individuals rarely map directly to the nomothetic description of change for the population and then only as a special case (i.e., ergodicity).

Increasingly, there is a recognition that it is inappropriate simply to treat the misalignment of individual and population characterizations as noise or error. In statistics, strategies from multilevel modeling have been applied to parse out the differences between intraindividual and interindividual effects and build models that structure the relationships between them (e.g., Berry & Willoughby, 2017; Curran & Bauer, 2011; Hamaker, Kuiper, & Grasman, 2015; Zhang, Browne, & Nesselroade. 2011). Such approaches take a step toward reconciling some of the tensions that exist between making sense of phenomena at the level of the individual and at the level of the group. However, they remain insufficient in their ability to construct the underlying data in ways that accommodate a wide range of idiographic influences on the quantities obtained for analysis. The chapters in this book present specific strategies for addressing this challenge.

Defining Measurement

Generally, there exist three major perspectives on what constitutes measurement, and each aligns with a different underlying ontological stance. While a full treatment of measurement theory is beyond the scope of this writing, it is necessary to understand how competing theories inform what it means to obtain measurement, how they might engage incommensurability

between idiographic and nomothetic facets of a phenomenon as a property of measurement, and under what assumptions about the nature of reality such integrated measurement can be achieved.

Representational measurement theory (Campbell, 1920; Krantz, Luce, Suppes, & Tversky, 1971; Michell, 2012) requires that measures meet several criteria: order, additivity, and homogeneity. This means that increments between scores must reflect exactly proportionate increments of the phenomenon measured, the sum of two scores must result in an accurate reflection of the joined instances of the phenomenon, and the nature of the phenomenon does not change in any way as a function of its score. These properties are often illustrated using measured length as an example: The scale unit (e.g., meter, inch, etc.) must appear the same number of times in two rods that appear to be the same length, the measured length of the two rods laid end to end must equal the sum of their respective lengths, and the meaning or nature of length cannot change as a function of how long or short a rod might be.

These representationalist standards of measurement entail a constructivist epistemology, in which the observed relationships between phenomena (and not the phenomena themselves) cause the relationship between quantities obtained through measurement; these observations, when appropriately quantified, give rise to measurement that reflects, but does not embody, the properties observed in the underlying phenomena. Accordingly, theoretical entities do not exist in a directly measurable or observable way and are merely human constructions used to make sense of obtained data.

Borsboom (2005) argues that this perspective entails two consequences, which bear directly on the relevance of incommensurability. First, representationalism does not have a ready account of measurement error, because the quantities obtained using true measures must directly reflect the observation of the thing measured, consistent with its measurement criteria. While probabilistic framing could be applied to mediate the relationship between attained measures and observed phenomena, it would entail the use of a latent construct, which would belie the tenets of representationalism. Second, a difference between two instances of a phenomenon can only be inferred through the difference in measured properties. Consequently, representationalism's effort to "construct quantitative metrics from qualitative observations" reduces to a reliance on a "quantitative metric directly at the level of the comparisons made" (p. 107). Thus, measurement cannot reflect differences that are not both observable and considered to be meaningful (i.e., worthy of counting by the observer) in the underlying phenomena. This cuts off the ability to utilize incommensurability, which is, by definition, information lost in an incomplete or imperfect mapping between an entity or relation, its qualitative observation, and the quantitative representation of that relation.

Classical test theory (CTT; Lord & Novick, 1968; Novick, 1966) directly accommodates the notion of measurement error by asserting that any obtained measurement reflects a true score (tau) coupled with measurement error. Further, measurement error is assumed to be randomly distributed, which implies that its influence on assessment of tau would be nonexistent over an infinite number of measurement instances. In a further divergence from representationalism, CTT asserts that measurement need not necessarily be on a continuous scale. Lord and Novick are explicit that categorical outcomes can likewise be understood as the representation of a true score and error.

Although there is a tendency to assume that the true score, tau, represents a realist ontological position, it is important to recall that true scores are only assessed within specific measures, rather than across various measures that would purport to measure the same underlying construct. Thus, a true score is not a property of the world, but instead, a property of a measure in practice. Conversely, CTT holds that test validity is established by a monotonic relationship with a hypothetical criterion variable that exists external to the measurement system. However, the number of hypothetical variables with which a given measure might correlate monotonically is unconstrained, so CTT cannot support claims of measurement unique to a specific construct or phenomenon. Accordingly, Borsboom (2005) characterizes the theory as operationalist, where it leverages robust mathematical strategies to generate a system for assessing the sufficiency of measures under fixed assumptions that ultimately preclude any specific ontological stance. Accordingly, CTT is accommodating of incommensurability, but it cannot engage its meaning beyond a departure from tau or randomly distributed error.

Latent variable theory (LVT; Borsboom, Mellenbergh, & van Heerden, 2003), on the other hand, takes a strong realist stance in its use of latent constructs as entities that exist independent of measurement. Whereas representationalism asserts that measurement scales represent observed (or hypothetically observed) relations between numeric systems and phenomena, LVT permits measures a direct, stochastic relationship with an extant latent variable that holds ontological status. Further, it avoids CTT's multiple correlation problem with validating measures (i.e., measures are valid to detect any and every construct with which they monotonically covary) by permitting parameterization of the latent variable to ensure unidimensionality within a measurement model. Further, polytomous item response theory (IRT) models have successfully characterized polytomous values as latent constructs. Models such as nominal response, partial credit, rating scale, and graded response (Ostini & Nering, 2006) evaluate relative quantities of intensity that do not necessarily reflect equal distances between sequential values (e.g., Likert scales).

From a realist position, latent constructs hold causal relationships with measures—their existence and properties (stochastically) cause the scores

generated through the act of measurement. Accordingly, the latent (i.e., unobserved) nature of existing phenomena is explicit in accommodating incommensurability between phenomena, observation of them, and identification of trends within observations.

Incommensurability as a Consequence of Latent Variable Theory's Realist Ontology

One of the fundamental measurement questions challenging LVT is the logical support for allowing continuous manifest variables to measure latent categorical variables and vice versa. This is especially difficult at face value when one adopts fundamental measurement criteria from the representationalist tradition, such as the need for concatenation of two measured values on a single scale to align with the properties of the concatenation of the things measured. For example, the sum of the measured lengths of two rods must equal that measured length of those rods laid end to end. However, when the nature of the underlying latent construct is categorical, it is not clear how the sum of two continuous manifest values would be reflected.

Within current latent variable modeling practices, latent means, variances, and other properties may differ within latent classes. In other words, each category of the latent categorical variable can reflect continuous properties within it. Further, because those properties are anchored to a specific class within the latent variable and all classes are both mutually exclusive and exhaustive (Litson, Thornhill, Geiser, Burns, & Servera, 2019), a concatenation of continuous values at the manifest level could result in a change in category membership in accordance with the parameters underlying each latent category.

Conversely, manifest categorical variables can viably measure a latent continuous variable in one of several ways. First, at a foundational level, dichotomous categorical variables can represent measurement units that yield continuous data when aggregated, an idea embodied in IRT (Guttman, 1950; Rasch, 1960). For example, when measuring length, we identify a fixed number of homogenous units to provide a value. Each unit can be considered dichotomous in value as it maps onto the segment to be measured. Accordingly, a length of 1 meter is the collective frequency of 100 dichotomous values for the categorical variable of a centimeter where categories are non-overlapping. While length measurements routinely include fractional values, these are never exact—they are rounded to the nearest unit of practical significance. Accordingly, any measured length is observed or represented as a finite number of non-zero dichotomous values, where the category is defined by the smallest quantity meaningful for a particular use. Whatever fractional length is left unaccounted for in the selection of the smallest meaningful unit reflects incommensurability between the measure and the phenomenon.

This account reflects the ontological differences between the assumptions of representational measurement and latent variable measurement theories. In the representationalist account, the measurement occurs as a mirror of the perceptions of phenomena. Accordingly, any difference in measurable quantities that is unnoticeable is absent. In contrast, the realist ontology of LVT accepts that the measured units are caused by the underlying phenomenon itself, in conjunction with other factors that could affect the stochastic ability of the measure to capture it fully. Thus, incommensurability is a construct fundamental to LVT that is not possible within the ontological constraints of representational measurement.

Importantly, the properties of a latent class are not exclusively quantitative. While statistical tools and principles are most typically applied in applications of LVT, there is nothing inherent to the theory that specifies only quantitative properties. Indeed, under a realist ontology, latent constructs exist independent of the tools used to gauge them. Accordingly, qualitative properties can manifest through any mode of observation suited to do so—which would include idiographic experiences of them.

Critical Realism and the Integration of Nomothetic and Idiographic Perspectives

Observation of underlying phenomena in LVT is stochastic in terms of the precision with which they influence the manifest (obtained) observations. It is also the case that other latent variables can influence the manifest value of a measure. Similarly, idiographic perspectives on an event or phenomenon can be influenced by multiple latent factors that can include physical vantage point, social positionality, and individual sense-making. Accordingly, these latent factors must be real in a fundamental sense to cause and influence these perceptions.

In IRT, latent influences may cause individual items to function differentially across subpopulations of interest (e.g., gender). Addressing such differential item function (DIF) typically leads to the exclusion of those items to enhance comparability of scores. However, DIF is not inherently problematic for measurement if its latent sources can be identified and incorporated into a larger causal account of the relations between latent constructs and their manifest values (Benítez & Pedilla, 2014; Hitchcock & Johanson, 2015).

Accordingly, LVT and its realist ontology can accommodate idiographic phenomena; it is only the extent to which such possibilities have been pursued and developed that are limited. To the extent that such incommensurability is influenced by other latent phenomena (e.g., social or physical positionality, individuals' prior knowledge or beliefs, etc.), these influences can be modeled as part of the measurement process. Indeed, in contrast to