MAIMON RESEARCH LLC

ARTIFICIAL INTELLIGENCE LARGE LANGUAGE MODEL INTERROGATION

REPRESENTATIONAL MEASUREMENT FAILURE IN HEALTH TECHNOLGY ASSESSMENT

AUSTRALIA: A MEASUREMENT-BASED CURRICULUM FOR HEALTH TECHNOLOGY ASSESSMENT

Paul C Langley PhD Adjunct Professor, College of Pharmacy, University of Minnesota, Minneapolis, MN

LOGIT WORKING PAPER No 1201 JUNE 2026

www.maimonresearch.com

Tucson AZ

ABSTRACT

Health technology assessment (HTA) has, for more than four decades, been dominated by a curriculum and research framework that gives priority to economic evaluation, utility measurement, simulation modelling and cost-effectiveness analysis. At the same time, there has been little systematic attention to the principles of measurement that underpin lawful quantitative claims throughout the sciences. Recent large language model (LLM) interrogations of PBAC and seven Australian HTA research centers suggest that this omission is not confined to research practice but is embedded within the curriculum knowledge base itself. Concepts central to measurement science, including scales of measurement, representational measurement, unidimensionality, manifest and latent attributes, ratio measurement and falsification, are either absent or only weakly represented.

This paper argues that the persistence of measurement inversion in HTA is a direct consequence of these curriculum deficiencies. The curriculum knowledge base is defined as the collection of concepts, principles, educational resources and research outputs that shape the intellectual environment within which future HTA practitioners are trained. If students are not exposed to the principles necessary to evaluate quantitative claims, then the continued acceptance of measurement inversion should not be surprising. Curriculum content therefore becomes the critical link between measurement theory and research practice.

To address these deficiencies, the paper proposes a reconstructed curriculum framework organized around six integrated domains: measurement foundations; manifest attributes and therapy assessment; Rasch measurement and latent attributes; evaluation of latent attributes; evaluation of quantitative claims; and scientific assessment and the evolution of objective knowledge. The framework begins with attributes, empirical structures and representational measurement before progressing to the assessment of manifest and latent attributes, the construction of Rasch logit ratio scales, the evaluation of therapy impact claims and the standards of normal science. Particular emphasis is placed on the distinction between scores and measures, the limitations of ordinal summation and composite indices, the requirements of admissible transformations and dimensional homogeneity, and the role of falsification in the growth of objective knowledge.

The proposed curriculum represents a fundamental departure from conventional HTA education. Rather than training students to apply established analytical frameworks, it seeks to equip them with the conceptual tools necessary to determine whether those frameworks satisfy the standards required for scientific inquiry. The paper concludes that meaningful reconstruction of HTA must begin with curriculum reconstruction. Unless students and researchers are exposed to the principles of measurement science and the standards of normal science, measurement inversion will continue to define both HTA education and HTA practice. The proposed curriculum provides a pathway toward a measurement-based HTA capable of supporting evaluable, replicable and falsifiable claims regarding therapy impact.

INTRODUCTION

The curriculum knowledge base is defined as the body of concepts, principles, methods and educational resources that prepare students and researchers to undertake the assessment of therapy impact. It extends beyond formal course descriptions to encompass teaching materials, methodological guidance, faculty publications, seminar programs, training workshops, doctoral supervision, research protocols and other sources through which disciplinary knowledge is transmitted. The curriculum knowledge base therefore represents the intellectual environment within which future HTA practitioners acquire their understanding of evidence, measurement and scientific inquiry.

For the purposes of this assessment, the curriculum knowledge base is not concerned with the technical details of specific analytical methods. Rather, the focus is on the foundational concepts that support quantitative claims. The central question is straightforward: are students and researchers in HTA exposed to the principles necessary to understand the measurement requirements of therapy assessment? This question is important because the validity of any quantitative claim depends upon the measurement properties of the quantities being analyzed. Without an understanding of measurement, there is no basis for determining whether numerical operations and subsequent conclusions are legitimate.

The curriculum knowledge base examined in this assessment is therefore organized around a series of foundational concepts. These begin with the identification of an attribute as the object of measurement. Every therapy assessment requires specification of an outcome or attribute of interest. Before a claim can be advanced, the attribute must be identified and its measurement status established. This introduces the requirement that students and researchers be familiar with the principal scales of measurement and the properties that distinguish nominal, ordinal, interval and ratio scales.

A second component of the curriculum knowledge base concerns representational measurement. Students and researchers should be aware that quantitative claims are governed by a set of measurement principles and axioms that determine whether arithmetic operations are admissible. These concepts are fundamental to all scientific disciplines that rely upon quantitative evidence. Closely related is the requirement that attributes be demonstrated to be unidimensional before measurement is possible. Without a single underlying attribute there is no basis for claiming that numerical observations represent different magnitudes of the same quantity.

The curriculum knowledge base must also address the distinction between manifest and latent attributes. Manifest attributes are directly observable and may often support linear ratio measurement. Latent attributes, by contrast, are not directly observable and require a measurement model to estimate possession of the attribute. Understanding this distinction is essential because the two classes of attributes require different approaches to measurement. Awareness of these requirements provides the foundation for lawful quantitative claims regarding therapy impact.

Finally, the curriculum knowledge base encompasses the scientific standards that govern claims assessment. Therapy impact claims should be capable of empirical evaluation and potential falsification. The purpose of measurement is not merely to generate numbers but to support claims that can be tested, replicated and challenged through observation.

Taken together, these concepts define the minimum educational foundations required for a measurement-based approach to HTA. The curriculum knowledge base therefore serves as the critical link between measurement theory and research practice. If these concepts are absent from educational programs, then their absence from subsequent HTA research and policy should not be unexpected. Curriculum content is thus not a peripheral concern. It is the primary mechanism through which the intellectual foundations of a discipline are established, maintained and reproduced.

THE FOUNDATIONS OF A RECONSTRUCTED HTA CURRICULUM

The curriculum interrogations demonstrate that a number of concepts central to scientific assessment are either absent or only weakly represented within contemporary HTA education. If HTA is to support lawful quantitative claims regarding therapy impact, these deficiencies must be addressed directly. The six curriculum domains outlined below represent the minimum educational foundations required for a measurement-based approach to therapy assessment. Together they provide a coherent framework that links measurement theory, attribute assessment, Rasch measurement, claims evaluation and scientific falsification into an integrated curriculum for HTA education and professional development.

MODULE 1: MEASUREMENT FOUNDATIONS

This module provides the conceptual foundations upon which all quantitative claims in health technology assessment must rest. The starting point for any scientific assessment is not arithmetic, modelling or statistical analysis but the identification of the attribute of interest and the determination of its measurement properties. Before quantities can be added, multiplied, compared or incorporated into analytical frameworks, it must first be demonstrated that they satisfy the requirements of measurement.

The module introduces the distinction between empirical structures and numerical structures, emphasizing that measurement is the process by which empirical relationships are represented numerically. Students are introduced to manifest and latent attributes, the requirement for unidimensionality, the principal scales of measurement and the principles of representational measurement. Particular attention is given to admissible transformations, invariance, dimensional homogeneity and the conditions required to support quantitative claims.

The module concludes with the fundamental principle that measurement precedes arithmetic. This principle is recognized throughout the physical and social sciences but has received limited attention within contemporary HTA. Understanding these foundations is essential because every subsequent component of the curriculum, whether concerned with manifest attributes, latent attributes, Rasch measurement or therapy impact claims, depends upon the concepts introduced in this module. Without an understanding of measurement foundations, students may become proficient users of analytical methods while remaining unable to determine whether the quantities entering those analyses are lawful measures.

• **Attributes**

- Empirical Structures
- Numerical Structures
- Manifest Attributes
- Latent Attributes

• **Unidimensionality**

- The single attribute requirement
- Measurement versus classification

• **Scales of Measurement**

- Nominal Scales
- Ordinal Scales
- Interval Measurement
- Ratio Measurement
- Permissible Analyses

• **Representational Measurement**

- Measurement as Numerical Representation
- Axioms of Representational Measurement
- Quantitative Claims
- Admissible Transformations
- Invariance
- Dimensional Homogeneity

• **Measurement Precedes Arithmetic**

- Arithmetic as a consequence of measurement
- Conditions for addition, subtraction, multiplication and division
- Measurement versus numerical manipulation

MODULE 2: MANIFEST ATTRIBUTES AND THERAPY ASSESSMENT

This module examines the measurement and assessment of manifest attributes. These are attributes that can be observed directly and are therefore capable of supporting empirical observation and linear ratio measurement. Examples include survival, hospital admissions, treatment persistence, medication possession and healthcare resource utilization. Because manifest attributes possess observable empirical structures, they provide the most straightforward basis for therapy assessment. The emphasis of this module is not merely on measurement but on the development and evaluation of prospective therapy impact claims. Students are introduced to protocol-based assessment, the specification of success criteria and the requirements for evaluable, replicable and falsifiable claims. The module demonstrates how lawful ratio measurement provides a foundation for credible assessments of therapy impact in real-world settings.

• Manifest Attributes

- Directly observable attributes
- Empirical observation and measurement
- Observable versus inferred outcomes
- Selection of target attributes

• Linear Ratio Measurement

- Properties of ratio scales
- True zero
- Magnitude comparisons
- Multiplication and division
- Admissible arithmetic operations

• Manifest Attributes in HTA

- Survival
- Mortality
- Hospital admissions
- Hospital days
- Emergency department visits
- Physician visits
- Medication possession
- Treatment persistence
- Treatment switching
- Resource utilization

• Prospective Therapy Impact Claims

- Defining therapy impact claims
- Attribute-specific claims
- Timeframes for assessment
- Success and failure criteria
- Evaluability requirements

• Protocol Development

- Protocol design
- Target populations
- Comparator specification
- Measurement requirements
- Data collection standards

- Replication and reproducibility

• Assessing Manifest Attribute Claims

- Magnitude of change
- Clinical relevance
- Replication
- Falsification
- Reporting standards

MODULE 3: RASCH MEASUREMENT AND LATENT ATTRIBUTES

This module introduces Rasch measurement as the required framework for the development of measures for latent attributes. Unlike manifest attributes, latent attributes such as pain, fatigue, anxiety, depression, physical functioning and need fulfilment cannot be observed directly. Their existence must be inferred from observable manifestations. The challenge is therefore not merely to collect responses but to transform those responses into measures that satisfy the requirements of representational measurement.

The significance of Rasch measurement lies in its unique position within the measurement sciences. Rasch provides the only framework that is both necessary and sufficient for the construction of measures for latent attributes. It does so by creating a conjoint measurement structure in which persons and items are simultaneously calibrated on a common latent continuum. The resulting measure is independent of the particular sample of respondents and independent of the particular set of items used to estimate attribute possession, provided model requirements are met. This property of specific objectivity distinguishes Rasch measurement from conventional scoring systems and from statistical models that focus on prediction rather than measurement.

The module emphasizes that summed scores are not measures. Summed scores merely rank individuals and therefore possess, at best, ordinal properties. Rasch measurement was developed specifically to overcome the limitations of ordinal scores by constructing invariant measures that support meaningful comparisons of attribute possession. The objective is not to estimate changes in scores but to estimate changes in the extent to which individuals possess a latent attribute. Therapy assessment therefore focuses on changes in the distribution of attribute possession within a target population rather than changes in arbitrary numerical scores.

Successful completion of this module provides students with the conceptual and practical foundations required for latent attribute measurement and prepares them for the subsequent assessment of therapy impact claims involving latent attributes.

Rasch Measurement Theory

• Measurement versus Scoring
• Latent Variable Theory
• Conjoint Measurement
• Item-Person Interaction

• Specific Objectivity
• Invariance

The Rasch Model

• Person Ability and Item Difficulty
• Probabilistic Measurement
• Model Requirements
• Unidimensionality
• Local Independence

Rasch Logit Ratio Scales

• Logits as Measures
• Ratio Interpretation through Odds
• Calibration of Persons and Items
• Change in Attribute Possession

Instrument Development (I): Constructing the Scale

• Item Selection
• Respondent Selection
• The Rasch Matrix
• Iteration
• The Provisional Logit Ratio Scale

Instrument Development (II): Finalizing the Scale

• Fitting the Rasch Model
• Item Fit and Person Fit
• Differential Item Functioning
• Category Functioning
• The Final Rasch Logit Ratio Scale

Therapy Assessment and Latent Attributes

• Item Responses and Therapy Impact
• Latent Attribute Possession
• Measuring Change in Possession
• Distribution of Possession in a Population
• Average Possession and Therapy Impact Claims

MODULE 4: EVALUATING LATENT ATTRIBUTES

The assessment of latent attributes represents both the greatest challenge and the greatest opportunity facing the reconstruction of health technology assessment. Manifest attributes such as survival, hospital admissions and treatment persistence can be observed directly and measured through linear ratio scales. Latent attributes are fundamentally different. Attributes such as pain, fatigue, anxiety, depression, physical functioning and need fulfilment cannot be observed directly. Their existence must be inferred from observable manifestations. The challenge is therefore to construct measures that estimate the extent to which individuals possess the latent attribute and to determine whether therapy changes that possession.

Remarkably, despite more than four decades of methodological development, HTA has failed to exploit the opportunities offered by Rasch measurement. While thousands of patient-reported outcome instruments have been developed, none have been designed specifically to provide measures of latent attribute possession suitable for therapy assessment. Instead, the dominant tradition has relied upon summing item responses to create scores and then treating those scores as though they were measures. Rasch measurement was developed precisely to overcome this limitation. Its objective is not the construction of scores but the creation of invariant measures of latent attributes.

This situation is particularly surprising because the practical tools required to implement Rasch measurement have been available for decades. Software systems such as WINSTEPS, available since the 1980s, provide a comprehensive environment for Rasch analysis, scale development, calibration and validation. The methodological framework exists, the software exists and the measurement theory exists. What has been missing is a commitment to constructing instruments whose explicit purpose is to estimate latent attribute possession and evaluate therapy impact.

The process itself is straightforward. The starting point is the identification of a latent attribute of interest. Items are then proposed that reflect increasing levels of difficulty in possessing that attribute. Responses from an appropriate population are assembled into a Rasch matrix and subjected to iterative analysis. Through successive refinements, items that fail model expectations are removed or modified, producing a provisional logit scale that is subsequently tested for fit, invariance and stability. The outcome is a Rasch logit ratio scale capable of estimating both person measures and item calibrations on a common latent continuum.

The significance of this process extends beyond instrument development. Once a valid Rasch instrument has been constructed, the focus of therapy assessment changes fundamentally. The research question is no longer whether scores increase or decrease. The question becomes whether therapy changes the extent to which individuals possess the latent attribute and whether the distribution of attribute possession within the target population shifts over time. Average possession may increase, decrease or remain unchanged. These are evaluable and falsifiable claims that can be tested through observation and replication.

Without the development of Rasch-based instruments for latent attribute possession, the future of HTA assessment of latent outcomes is severely constrained. With them, HTA gains access to a scientifically defensible framework for evaluating therapy impact in areas that matter most to patients. The defining feature of the Rasch framework is that therapy impact is assessed through changes in the estimated possession of a latent attribute rather than through changes in summed scores. Conventional patient-reported outcome instruments typically assume that higher scores indicate greater levels of the attribute and that differences in scores reflect differences in therapy impact. This assumption is rarely justified. Summed scores are merely ordinal rankings. They indicate that one person has a higher score than another, but they do not establish the magnitude of the difference between those scores or whether equal score changes represent equal changes in the underlying attribute.

Rasch measurement approaches the problem differently. The objective is not to count positive responses but to evaluate the entire pattern of responses across all items in relation to the difficulty hierarchy of those items. Each item occupies a specific location on the latent continuum, representing a particular level of difficulty in possessing the attribute. Similarly, each respondent occupies a location on the same continuum, representing the estimated extent to which the attribute is possessed.

Possession is therefore inferred from the relationship between person measures and item calibrations. A respondent who endorses only the easiest items demonstrates a lower level of possession than a respondent who also endorses more difficult items. The critical point is that two individuals may have identical summed scores yet possess different levels of the attribute because the pattern of their responses across the item hierarchy differs. One respondent may endorse several difficult items while failing some easier items; another may endorse only easier items. The summed score treats these patterns as equivalent. Rasch measurement does not.

The same principle applies to therapy assessment. The question is not whether the number of positive responses increases following treatment. Rather, the question is whether the pattern of responses changes in a manner consistent with increased possession of the latent attribute. Improvement is reflected by movement along the latent continuum, demonstrated through successful endorsement of increasingly difficult manifestations of the attribute. The resulting change is expressed as a change in the estimated level of attribute possession rather than as a change in score.

Consequently, the focus of evaluation shifts from arithmetic comparisons of scores to measurement of possession. Therapy impact is assessed through changes in person measures, changes in the distribution of possession within the target population and changes in average possession across the population. This provides a direct and scientifically meaningful estimate of therapy impact that is grounded in the measurement of the latent attribute itself rather than in the manipulation of ordinal scores.

Nature of Latent Attributes

• What is a latent attribute?
• Manifest versus latent attributes
• Latent variables and empirical reality
• Why latent attributes cannot be observed directly

Identifying the Target Attribute

• Defining the attribute of interest
• Theoretical foundations of the attribute

• Attribute boundaries
• Distinguishing attributes from manifestations
• Attribute specification for therapy assessment

Unidimensionality

• The single attribute requirement
• Demonstrating unidimensionality
• Consequences of multidimensionality
• Unidimensionality as a prerequisite for measurement

Manifestations of an Attribute

• Observable indicators of possession
• Manifestations versus attributes
• Item development from manifestations
• Ordering manifestations by difficulty of possession

Attribute Possession

• The concept of possession
• Extent of possession
• Person measures and attribute possession
• Population distributions of possession

Measuring Therapy Impact

• Baseline possession of the attribute
• Post-treatment possession of the attribute
• Change in possession
• Distributional shifts in possession
• Average possession within a target population

Evaluating Claims

• Prospective claims for latent attributes
• Evaluable claims
• Replicable claims
• Falsifiable claims

Module 5: Evaluating Claims

The purpose of measurement is not to generate numbers but to support credible claims regarding therapy impact. Yet one of the most persistent weaknesses in contemporary HTA is the failure to distinguish between the existence of numerical outputs and the existence of valid measures. Numbers alone do not create evidence. Before any claim can be accepted, the measurement properties of the quantities involved must be established and the proposed analyses shown to be consistent with those properties.

This module focuses on the evaluation of quantitative claims and the standards that determine whether those claims are scientifically defensible. The starting point is the distinction between scores and measures. Scores are often created through the summation of responses to questionnaire items or through the aggregation of multiple indicators. Measures, by contrast, require demonstration that the attribute satisfies the requirements of measurement and that the resulting numerical representation supports the proposed analytical operations. Failure to recognize this distinction has contributed significantly to the acceptance of numerical outputs that possess little or no measurement significance.

Particular attention is given to ordinal summation and composite indices. Across HTA there is a tendency to assume that numerical aggregation transforms observations into measures. This assumption is mistaken. Adding ordinal responses cannot create interval or ratio properties. Similarly, combining multiple attributes into a composite index does not establish that the resulting score represents a measurable quantity. Students are encouraged to evaluate whether an index represents a genuine measure or merely a convenient numerical summary.

Particular attention is also given to composite indices, which are often presented as measures despite combining multiple heterogeneous attributes into a single numerical score. The module demonstrates why aggregation alone cannot create measurement and why composite indices typically fail the requirements of unidimensionality, dimensional homogeneity and representational measurement.

The module also introduces the concept of admissible transformations. Different scales of measurement support different forms of mathematical manipulation. The validity of arithmetic operations depends upon the properties of the scale. A quantitative claim is therefore only as credible as the measurement structure that supports it. Understanding admissible transformations is essential for determining whether addition, subtraction, multiplication or division are scientifically legitimate.

Finally, the module addresses standards of evidence. Quantitative claims must be supported by lawful measurement, transparent analytical procedures and clearly specified evaluation criteria. Numerical sophistication cannot compensate for the absence of measurement. The central question is always the same: does the claim rest upon quantities that possess the measurement properties required for the proposed analysis?

The objective of this module is therefore to equip students with the ability to evaluate claims rather than simply accept them. By distinguishing scores from measures, identifying the limitations of ordinal summation and composite indices, understanding admissible transformations and applying appropriate evidence standards, students acquire the skills necessary to determine whether a proposed therapy impact claim is scientifically credible. This capability forms a critical bridge between measurement theory and the practical assessment of evidence in HTA.

Scores versus Measures

• The distinction between scores and measures
• Numerical rankings versus measurement
• Why scores are not measures
• Implications for therapy assessment

Ordinal Summation

• Construction of summed scores
• Ordinal properties of summed responses
• Limitations of ordinal arithmetic
• Why summation cannot create measurement

Composite Indices

• Construction of composite scores
• Aggregation of heterogeneous attributes
• Absence of a common underlying attribute
• Failure of unidimensionality
• Lack of dimensional homogeneity
• Why aggregation cannot create measurement
• Composite scores as numerical constructions
• Implications for therapy assessment

Admissible Transformations

• Scale properties and transformations
• Nominal scales
• Ordinal scales
• Interval scales
• Ratio scales
• Admissible arithmetic operations

Quantitative Claims

• Specification of therapy impact claims
• Measurement requirements for claims
• Manifest attribute claims
• Latent attribute claims
• Evaluability requirements

Evidence Standards

• Measurement as a prerequisite for evidence
• Replication and reproduction
• Falsification

• Protocol assessment
• Reporting standards

Evaluating Claims in HTA

• Are the quantities measurable?
• Are the arithmetic operations admissible?
• Is the claim dimensionally homogeneous?
• Does the claim rest upon a measurable attribute?
• Can the claim be evaluated empirically?
• Can the claim be replicated?
• Can the claim be falsified?

Module 5 becomes the quality-control module. Everything that has been learned in the first four modules is now brought together and applied to the evaluation of claims. The central question becomes: Does this claim rest on lawful measurement and does it satisfy the standards required for scientific evidence?

MODULE 6: SCIENTIFIC ASSESSMENT AND THE EVOLUTION OF OBJECTIVE KNOWLEDGE

The final module addresses a question that is rarely considered in contemporary HTA but is central to scientific inquiry: has the assessment contributed to the evolution of objective knowledge? The purpose of measurement, evidence generation and therapy assessment is not simply to produce numerical outputs. It is to advance understanding through the formulation and testing of empirical claims. Scientific progress occurs when claims are exposed to observation, challenged through replication and potentially rejected through falsification. Knowledge advances because claims are vulnerable to failure.

This principle lies at the heart of normal science. A scientific claim must be capable of empirical evaluation. The claim must be specified in advance, the conditions for success and failure identified and the results open to independent replication. Without these requirements there is no mechanism through which objective knowledge can evolve. Numbers alone are not evidence. Numerical sophistication does not substitute for scientific inquiry.

The importance of this distinction becomes apparent when considering the reference case framework that dominates contemporary HTA. The reference case does not generate evaluable claims regarding therapy impact. Instead, it generates assumption-driven simulations of hypothetical future events. The process begins with preference scores, proceeds through utility algorithms, creates quality-adjusted life years and culminates in cost-effectiveness claims derived from simulation models. At every stage the analysis depends upon assumptions rather than observation. The final result is not a claim regarding a measurable outcome but an estimate of what might occur under a specified set of assumptions.

The problem is not merely that the pathway from preferences to utilities, from utilities to QALYs and from QALYs to cost-effectiveness claims fails the requirements of measurement. The deeper problem is that the resulting claim cannot contribute to the growth of objective knowledge. The simulated lifetime outcome does not exist and cannot be observed. There is no protocol through which the claim can be replicated. There is no empirical test capable of demonstrating that the claim is false. The result is an imaginary numerical construction rather than a scientific proposition.

By contrast, a measurement-based approach to HTA focuses on prospective claims regarding manifest and latent attributes. These claims are specified in advance, measured through lawful instruments and assessed within defined time horizons. Success and failure criteria are explicit. Replication is possible. Falsification is possible. The outcome of the assessment therefore contributes directly to the accumulation of objective knowledge.

The purpose of this module is to place HTA within the broader philosophy of science. Students are introduced to the concepts of evaluability, replication, reproduction and falsification and are encouraged to view therapy assessment as part of an ongoing process of knowledge generation rather than numerical modelling. The central question is not whether a model can be constructed, but whether the resulting claim contributes to objective knowledge. If a claim cannot be tested, replicated or potentially refuted, it lies outside the boundaries of normal science and cannot provide a foundation for scientific decision making.

• Normal Science

- Scientific inquiry
- Empirical claims
- Objective knowledge
- Measurement and science

• Prospective Claims Assessment

- Claim specification
- Success and failure criteria
- Protocol assessment
- Real-world evaluation

• Replication and Reproduction

- Independent verification
- Reproducibility
- Generalizability
- Evidence accumulation

• Falsification

- Testable claims
- Conditions for refutation
- Scientific progress

- Evolution of knowledge

• Numerical Storytelling

- Assumption-driven models
- Imaginary future states
- Simulation versus observation
- Limitations of the reference case

• Scientific Standards for HTA

- Evaluable claims
- Measurable outcomes
- Replicable evidence
- Falsifiable claims
- Contribution to objective knowledge

CONCLUSION

For more than 40 years HTA has been locked into a curriculum and research framework whose primary function has been to support and legitimize the reference case. In doing so, the discipline has largely insulated itself from the standards that govern quantitative claims throughout the physical and social sciences. The result has been a remarkable intellectual inversion. Rather than beginning with measurement and proceeding to arithmetic, modelling and claims assessment, HTA has accepted numerical constructions as evidence without first establishing whether the quantities involved satisfy the requirements of measurement. Generations of faculty, researchers and students have inherited this framework and, in many cases, accepted its assumptions without question.

The curriculum assessment presented in this paper suggests that this outcome is not accidental. Across PBAC and the seven Australian HTA research centers there is little evidence that students and researchers are systematically exposed to the concepts necessary to evaluate quantitative claims. Scales of measurement, representational measurement, unidimensionality, manifest and latent attributes, ratio measurement and the standards governing scientific claims occupy, at best, a peripheral position within the curriculum knowledge base. Yet these are precisely the concepts required to determine whether a quantitative claim is lawful. Their absence provides a plausible explanation for the persistence of measurement inversion throughout HTA.

The significance of these findings extends beyond Australia. They raise a broader question concerning the purpose of HTA education itself. Is the objective to train analysts capable of implementing the reference case, or is it to educate scientists capable of evaluating whether the reference case satisfies the standards of measurement and scientific inquiry? These are fundamentally different educational objectives. The former reproduces an established methodology. The latter encourages critical evaluation of the foundations upon which that methodology rests.

The curriculum framework proposed in this paper is therefore unique in contemporary HTA. It begins not with economic evaluation, simulation modelling or cost-effectiveness analysis, but with measurement itself. Attributes, empirical structures, scales of measurement, representational measurement, manifest and latent attributes, Rasch measurement, evaluable claims and falsification become the organizing principles of the curriculum. The objective is to return HTA to the standards of normal science, where measurement precedes arithmetic and where empirical claims are judged by their ability to survive observation, replication and potential refutation.

The challenge facing HTA is not one of incremental reform. It is a question of intellectual reconstruction. A curriculum that neglects measurement cannot support a scientific discipline devoted to quantitative claims regarding therapy impact. Until these deficiencies are addressed, measurement inversion will continue to define both HTA education and HTA practice. The framework outlined here offers a different future: one in which therapy assessment is grounded in lawful measurement, prospective claims, empirical evaluation and the continuing evolution of objective knowledge. Whether HTA chooses to embrace that future remains an open question, but the curriculum foundations required to support it are now clear.

ACKNOWLEDGEMENT

I acknowledge that I have used OpenAI technologies, including the large language model, to assist in the development of this work. All final decisions, interpretations, and responsibilities for the content rest solely with me.