MAIMON RESEARCH LLC

ARTIFICIAL INTELLIGENCE LARGE LANGUAGE MODEL INTERROGATION

REPRESENTATIONAL MEASUREMENT FAILURE IN HEALTH TECHNOLGY ASSESSMENT

AUSTRALIA: MEASUREMENT INVERSION UNDERPINS HTA RESEARCH

The Neglect of Ratio Measurement

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

LOGIT WORKING PAPER No 733 JUNE 2026

www.maimonresearch.com

Tucson AZ

ABSTRACT

Health technology assessment (HTA) is concerned with claims regarding the impact of therapies on patients, healthcare systems and society. Whether the objective is to evaluate survival, hospital admissions, symptom burden, physical functioning or quality of life, all claims share a common requirement: they must be based upon valid measures capable of supporting empirical evaluation, replication and falsification. Yet the principles that govern measurement receive surprisingly little attention in many HTA programs. Students are routinely introduced to utilities, quality-adjusted life years (QALYs), cost-effectiveness models and simulation techniques without first being exposed to the measurement foundations required to determine whether the quantities employed are lawful measures.

This paper outlines the minimum measurement competencies that should be present in HTA curricula concerned with health outcomes research. The discussion begins with the foundations of measurement, including Stevens' scales of measurement and the axioms of representational measurement. Particular emphasis is placed on the principle that measurement precedes arithmetic and that quantitative claims depend upon the measurement properties of the quantities involved.

A distinction is then drawn between manifest and latent attributes. Manifest attributes, such as survival, hospital admissions and medication utilization, are directly observable and require unidimensional linear ratio measures characterized by a true zero, ratio comparisons and admissible arithmetic operations. Latent attributes, including pain, fatigue, depression, physical functioning and need fulfilment, cannot be observed directly and require a different measurement framework.

The paper argues that summed questionnaire scores are ordinal rankings rather than measures and that the measurement of latent attributes requires Rasch measurement theory. Rasch instruments are designed not to generate scores but to estimate the extent to which individuals possess a latent attribute. The outcome of interest is therefore not score change but change in the possession of the latent trait, assessed through shifts in the distribution of possession within the target population.

Finally, the paper considers the role of measurement in evaluating claims for therapy impact. Claims must be linked to clearly specified attributes, measured according to the standards appropriate to those attributes and subjected to empirical testing. The interrogation of over 230 HTA knowledge bases suggests that these foundations, particularly Rasch measurement theory, are largely absent from contemporary HTA education. The conclusion is straightforward: if HTA graduates are to evaluate therapy claims according to the standards of scientific inquiry, then ratio measurement for manifest attributes and Rasch logit ratio measurement for latent attributes should be regarded as essential foundations of their professional education.

INTRODUCTION

The purpose of health technology assessment (HTA) is to support claims regarding the impact of therapies on patients, healthcare systems and society. Whether the objective is to evaluate survival, symptom burden, functional status, healthcare resource utilization or quality of life, the underlying requirement is the same: the claims advanced must be capable of empirical evaluation. This requirement is so fundamental that it is often overlooked. Before a claim can be tested, replicated or falsified, it must first be expressed in terms of a valid measure. Measurement therefore precedes arithmetic, statistical analysis and modelling. Without measurement there can be no meaningful quantitative claim.

Surprisingly, the role of measurement receives little attention in many HTA programs that teach health outcomes research.. Students are routinely introduced to utilities, quality-adjusted life years (QALYs), cost-effectiveness models and simulation techniques, yet they may receive little formal instruction in the principles that determine whether a quantity is a measure in the first place. The consequence is that numerical outputs are often accepted as though they were measures, while the standards required to support that conclusion remain unexplored.

This omission is important because measurement is not merely a statistical exercise. It is governed by well-established principles developed across the physical and social sciences. The scales of measurement described by Stevens, together with the axioms of representational measurement, define the conditions under which arithmetic operations are permissible [1] [2] . Numbers can be assigned to observations in many ways, but not every numerical assignment constitutes a measure. The distinctions between scores, indices and measures are therefore critical. A quantity may be useful, familiar and widely accepted while still failing the requirements necessary to support quantitative claims.

The challenge facing HTA education is therefore straightforward. Students should be able to distinguish between directly observable attributes, such as survival or hospital admissions, and latent attributes, such as pain, fatigue or depression. They should understand that different classes of attributes require different approaches to measurement. They should also understand why arithmetic operations, including addition, multiplication and aggregation, depend upon the measurement properties of the quantities involved. Without this knowledge, students cannot critically evaluate claims regarding therapy impact, regardless of how sophisticated the associated statistical analyses may appear.

The objective of this paper is to outline the minimum measurement competencies required for the assessment of therapy impact. It provides an introduction to the foundations of measurement, the distinction between manifest and latent attributes, the requirements for ratio measurement and the role of Rasch measurement theory [3] [4] [5] .

. These concepts are not advanced topics. Rather, they represent the essential foundations that every HTA graduate should possess if claims for therapy impact are to meet the standards of scientific evaluation, replication and falsification.

THE FOUNDATIONS OF MEASUREMENT

Measurement occupies a unique position in science. It provides the bridge between observation and quantitative claims, allowing phenomena to be described, compared and evaluated in a systematic manner. In the absence of measurement, scientific inquiry is restricted to qualitative description and subjective interpretation. For this reason, every quantitative discipline, from physics and chemistry to psychology and economics, depends upon an understanding of the conditions that must be satisfied before numerical claims can be accepted as meaningful.

The starting point for measurement is the recognition that not all numbers are the same. Numbers can be assigned to observations for many reasons, but the assignment of numbers does not automatically create a measure. Telephone numbers, zip codes and jersey numbers all employ numbers, yet none are measures. The fact that numbers are present does not imply that arithmetic operations involving those numbers are meaningful. The question is not whether numbers exist, but whether those numbers possess the properties required to support quantitative comparisons and arithmetic operations.

A useful introduction to this issue is provided by Stevens' classification of scales of measurement. Stevens identified four broad classes of scales: nominal, ordinal, interval and ratio. Nominal scales simply classify observations into categories. Examples include blood type, gender or disease classification. Ordinal scales establish rank order, allowing observations to be placed in sequence from lower to higher levels of an attribute. Many questionnaire responses and severity classifications are ordinal because they indicate order without establishing equal distances between categories.

Interval scales represent a more demanding standard. In addition to rank order, equal intervals are assumed between adjacent points on the scale. Temperature measured in degrees Celsius is the traditional example. A difference of ten degrees has the same meaning throughout the scale. However, interval scales do not possess a true zero. Zero degrees Celsius does not represent the absence of temperature. Consequently, while addition and subtraction are permissible, ratio comparisons are not. Forty degrees Celsius is not twice as hot as twenty degrees Celsius.

Ratio scales represent the highest level of measurement. In addition to equal intervals, they possess a true, non-arbitrary zero that represents the absence of the attribute being measured. Examples include height, weight, distance and time. Ratio scales permit the full range of arithmetic operations, including multiplication, division and ratio comparisons. A person who weighs eighty kilograms weighs twice as much as a person who weighs forty kilograms. Such statements are meaningful because the scale possesses a true zero and equal units throughout its range.

The importance of Stevens' framework is that it links the properties of a scale to the arithmetic operations that can be performed. Arithmetic is not independent of measurement. The admissibility of arithmetic depends entirely upon the characteristics of the underlying scale. Multiplication and division require ratio measures. If a quantity lacks ratio properties, then ratio operations are not justified regardless of how sophisticated the subsequent statistical analysis may be.

While Stevens' classification provides a useful foundation, it applies primarily to directly observable or manifest attributes. The further development of measurement theory introduced a more formal set of requirements known as the axioms of representational measurement. These axioms address issues such as order, unidimensionality, additivity, invariance and admissible transformations. Together they provide the framework that determines whether numerical assignments can legitimately be interpreted as measures of an attribute.

For students of health technology assessment, the central lesson is straightforward. Measurement is not created by arithmetic, statistics or modelling. A quantity must first satisfy the requirements of measurement before arithmetic operations can be justified. This principle can be summarized in a single phrase: measurement precedes arithmetic. Every claim regarding therapy impact must therefore begin with a simple question: what exactly is being measured, and does that quantity satisfy the requirements necessary to support the arithmetic operations that follow? Without an answer to that question, numerical outputs remain numbers rather than measures, regardless of how frequently they appear in research reports, economic evaluations or policy decisions.

MEASUREMENT INVERSION: THE ABSENCE OF RATIO SCALES

Measurement inversion occurs when arithmetic is undertaken before the requirements of measurement have been established. Instead of first determining whether a quantity satisfies the conditions necessary for measurement, numerical operations are applied to scores, indices and preference values as though they were lawful measures. The result is a reversal of the scientific process. Measurement no longer precedes arithmetic; arithmetic is assumed to create measurement.

The central issue is the neglect of ratio scales. In both the physical and social sciences, multiplication, division and ratio comparisons require quantities measured on ratio scales. A ratio scale possesses a true zero, equal units throughout its range and supports meaningful statements regarding relative magnitude. Without ratio-scale properties, arithmetic operations that depend upon ratios are inadmissible. This principle is not a methodological preference. It is a foundational requirement of representational measurement.

The implications for health technology assessment are profound. Utilities, QALYs and cost-effectiveness claims all depend upon arithmetic operations that require ratio measures. Yet the knowledge-base interrogations of Australian HTA research centers, academic institutions and policy organizations indicate that these requirements are largely absent from the analytical frameworks that support teaching and research. Concepts such as ratio measurement, dimensional homogeneity, admissible arithmetic, unidimensionality and measurement preceding arithmetic receive little endorsement, while utilities, QALYs and simulation models are accepted as though their measurement foundations were self-evident.

This pattern is the hallmark of measurement inversion. Quantitative claims are advanced without first establishing the measurement properties necessary to support them. The consequence is a research environment in which numerical constructions are routinely treated as measures and arithmetic is substituted for measurement. The Australian interrogations suggest that this is not an isolated oversight but a pervasive feature of contemporary HTA. The neglect of ratio measurement therefore represents more than a methodological weakness. It calls into question the foundations upon which claims for therapy impact are constructed and evaluated.

MANIFEST AND LATENT ATTRIBUTES

The first step in evaluating any claim for therapy impact is to identify the attribute of interest. This question is more important than it may appear. Before a measure can be constructed, before arithmetic can be applied and before statistical analysis can be undertaken, it is necessary to know exactly what is being measured. Failure to distinguish between different classes of attributes is one of the principal reasons why confusion arises in health technology assessment and outcomes research.

Attributes fall into two broad categories: manifest attributes and latent attributes. The distinction is fundamental because each category requires a different approach to measurement.

Manifest attributes are directly observable. Their presence and magnitude can be determined without relying on subjective interpretation. Examples relevant to healthcare include survival time, hospital admissions, emergency department visits, length of stay, medication doses, physician visits and laboratory values. These attributes exist independently of any questionnaire or scoring system. They can be counted, timed, weighed or otherwise observed directly. The measurement challenge for manifest attributes is therefore relatively straightforward. The objective is to construct a scale that accurately represents the magnitude of the attribute and supports meaningful quantitative comparisons.

Many manifest attributes naturally possess ratio-scale properties. Time, distance, weight and counts of events all have a true zero and permit ratio comparisons. If a patient experiences four hospital admissions while another experiences two, the first patient has experienced twice as many admissions. If one treatment extends survival by twelve months while another extends survival by six months, the gain is twice as large. These statements are meaningful because the underlying measures possess ratio-scale properties.

Latent attributes are fundamentally different. They cannot be observed directly. Instead, they are inferred from behavior, responses or other indicators. Examples include pain, fatigue, anxiety, depression, physical functioning, social participation, quality of life and need fulfilment. Unlike hospital admissions or survival time, these attributes cannot be counted or weighed. Their existence is inferred from evidence rather than observed directly.

The measurement of latent attributes presents a more demanding challenge. Because the attribute itself cannot be observed, researchers must rely upon indicators, typically responses to carefully designed questionnaires. The difficulty is that responses are not the attribute. A patient's answer to a pain question is evidence regarding pain, but it is not pain itself. Similarly, a response indicating severe fatigue is evidence regarding fatigue, not a direct measure of fatigue. The task of measurement is therefore to transform these observations into a representation of the underlying latent attribute.

Historically, many health status instruments have attempted to address this problem by summing questionnaire responses to create total scores. While such scores may provide useful descriptive information, they do not automatically constitute measures. A numerical score is not equivalent to measurement. The crucial question is whether the score can be shown to represent a single underlying attribute with the properties required for quantitative comparisons. Without such evidence, the resulting number remains a score rather than a measure. The answer is Rasch measurement as the necessary and sufficient technique for transforming observations to ratio measurement .

The distinction between manifest and latent attributes is particularly important in pharmacy education because claims for therapy impact frequently involve both categories. A treatment may reduce hospital admissions, a manifest attribute, while also improving pain or functioning, which are latent attributes. The same measurement framework cannot be assumed to apply to both. Directly observable outcomes require one approach to measurement, while latent outcomes require another.

Recognition of this distinction immediately simplifies the measurement problem. Rather than treating all outcomes as though they were variations of the same concept, attention can be focused on the nature of the attribute itself. Is the attribute directly observable or indirectly inferred? Does it possess a natural zero? Can it be counted or timed? Or must it be estimated from responses to indicators? Once these questions are answered, the appropriate measurement framework becomes clear.

For students and faculty involved in health technology assessment, this distinction is essential. The validity of any claim regarding therapy impact depends not only on the outcome reported but also on whether the outcome has been measured appropriately. Understanding the difference between manifest and latent attributes therefore provides the foundation for all subsequent discussions of ratio measurement, Rasch measurement and the evaluation of quantitative claims. Before deciding how an attribute should be measured, it is first necessary to determine what type of attribute it is.

RATIO MEASUREMENT FOR MANIFEST ATTRIBUTES

Once an attribute has been identified as manifest, the measurement problem becomes considerably simpler. Manifest attributes are directly observable and can be counted, timed, weighed or otherwise recorded without relying upon inference from responses or indicators. Examples include survival time, hospital admissions, emergency department visits, medication utilization, physician consultations, intensive care unit days and laboratory values. Because these attributes are directly observable, the challenge is not to infer their existence but to measure their magnitude accurately. Before this can be accomplished, however, a number of requirements must be satisfied.

For a manifest attribute to support claims regarding therapy impact, six conditions should be met:

1. The attribute must be clearly specified.
2. The attribute must be unidimensional.
3. The scale must possess a true zero.
4. The scale must support ratio comparisons.

5. Arithmetic operations must be admissible.
6. Claims based on the measure must be evaluable and falsifiable.

These requirements are not arbitrary. Together they define the conditions under which a quantity can be regarded as a lawful measure capable of supporting scientific claims.

The first requirement is that the attribute must be clearly specified. Measurement begins with an explicit statement regarding what is to be measured. If the attribute is not clearly identified, there is no basis for constructing a measure. A claim that a therapy improves "healthcare utilization" illustrates the problem. Healthcare utilization may refer to hospital admissions, emergency department visits, physician consultations, diagnostic procedures or medication use. Each represents a different attribute. Unless the attribute is clearly defined, the meaning of the measure remains uncertain and any subsequent arithmetic becomes difficult to interpret.

The second requirement is unidimensionality. The measure must represent one and only one attribute. This principle lies at the heart of representational measurement. A measure of survival must measure survival. A measure of hospital admissions must measure hospital admissions. A measure that combines admissions, physician visits and medication use is no longer measuring a single attribute. Instead, it becomes a composite index. Composite indices may be useful descriptive tools, but they cannot automatically be assumed to possess the properties required for measurement. Before arithmetic operations can be justified, it must first be demonstrated that a single attribute is being represented.

The importance of unidimensionality is often underestimated. It is impossible to measure more than one attribute with a single measure. Consider a simple example. A researcher combines hospital admissions, emergency department visits and physician consultations into a single score intended to represent healthcare utilization. Although the resulting score contains numbers, it does not represent a single attribute. It is therefore impossible to determine exactly what the score measures. The presence of numbers does not create measurement. Measurement begins with the identification of a unidimensional attribute.

The third requirement is the presence of a true, non-arbitrary zero. Ratio measures differ from interval measures because zero represents the complete absence of the attribute being measured. Zero hospital admissions mean no admissions. Zero physician visits mean no visits. Zero days of hospitalization means no hospitalization. A meaningful zero point is essential because it permits ratio comparisons. Without a true zero, multiplication and division lose their substantive interpretation.

The fourth requirement follows directly from the third. The measure must support ratio comparisons. Ratio comparisons allow one observation to be interpreted relative to another. A patient experiencing four hospital admissions has experienced twice as many admissions as a patient experiencing two. A patient hospitalized for ten days has spent twice as long in hospital as a patient hospitalized for five days. Such statements are meaningful because the underlying scale possesses a true zero and equal units throughout its range.

The fifth requirement concerns admissible arithmetic. Arithmetic operations are not independent of measurement. The properties of the scale determine which operations are permissible. Addition and subtraction may be appropriate for some scales, while multiplication and division require ratio measures. This principle is often overlooked in health outcomes research and HTA. The fact that a numerical value exists does not imply that all arithmetic operations are justified. Arithmetic must be consistent with the measurement properties of the quantity involved. Measurement precedes arithmetic, not the reverse.

Examples of ratio measures are common throughout healthcare. Survival time, hospital admissions, days of hospitalization, medication doses, treatment persistence and emergency department visits can all be expressed as ratio measures provided they refer to a single attribute. Importantly, these measures do not require preference weights, utility algorithms or complex transformations. Their meaning derives directly from the attribute itself. The measure exists because the attribute is observable and the scale possesses the properties required for quantitative comparison.

The sixth requirement is that claims based on the measure must be capable of empirical evaluation and falsification. This requirement reflects the broader philosophy of science. A measure is valuable only to the extent that it supports claims that can be tested. Suppose a manufacturer claims that a therapy reduces hospital admissions by 25 percent over a twelve-month period. Because hospital admissions are measured on a unidimensional ratio scale, the claim can be evaluated. Data can be collected, comparisons can be made and the claim can be either supported or rejected. The measure therefore provides a foundation for replication and falsification.

The ability to support evaluable claims is perhaps the most important contribution of ratio measurement. Scientific progress depends upon the continuous testing of claims against observation. Measures that fail to support empirical evaluation contribute little to scientific understanding regardless of their complexity or popularity. In contrast, ratio measures permit straightforward comparisons between competing claims and provide a transparent basis for decision making.

For students of pharmacy and health technology assessment, the implications are straightforward. Measurement is not created by statistical analysis, econometric modelling or simulation exercises. Measurement begins with the identification of a clearly specified and unidimensional attribute. A scale must then be constructed that possesses a true zero, supports ratio comparisons and permits admissible arithmetic operations. Only when these conditions have been satisfied can claims regarding therapy impact be subjected to empirical evaluation and falsification.

Manifest attributes therefore provide the most direct route to the scientific assessment of therapy impact. They allow outcomes to be expressed in terms of observable quantities measured on ratio scales and evaluated through observation. The situation becomes more challenging when the attribute of interest cannot be observed directly. In such circumstances a different measurement framework is required. The challenge is no longer to measure a directly observable quantity but to estimate the extent to which a person possesses an underlying latent attribute. It is this problem that gives rise to Rasch measurement theory and the measurement of latent traits.

RASCH RATIO MEASUREMENT FOR LATENT ATTRIBUTES

The measurement of latent attributes presents a challenge that differs fundamentally from the measurement of manifest attributes. Manifest attributes such as survival, hospital admissions and medication use can be observed directly. Their magnitude can be counted, timed, weighed or otherwise recorded. Latent attributes are different. Pain cannot be observed directly. Neither can fatigue, anxiety, depression, physical functioning, social participation or need fulfilment. Their existence is inferred from evidence rather than observed directly. The problem facing measurement is therefore not how to count the attribute but how to construct a measure that accurately represents the extent to which an individual possesses that attribute.

The starting point is exactly the same as for manifest attributes. Before measurement can occur, the attribute must be clearly specified and demonstrated to be unidimensional. Unless a single attribute has been identified, there is nothing to measure. A measure of pain must measure pain and nothing else. A measure of fatigue must measure fatigue and nothing else. A questionnaire that combines multiple attributes into a single score does not create a measure. Instead, it creates a numerical summary whose interpretation is uncertain. Just as unidimensionality is a prerequisite for ratio measurement of manifest attributes, it is also a prerequisite for the measurement of latent attributes.

Historically, many attempts to assess latent attributes have relied upon questionnaires composed of multiple items. Respondents answer questions and the responses are summed to create a total score. This approach remains common throughout health outcomes research and health technology assessment. The assumption is that larger scores represent greater amounts of the attribute and that changes in scores represent changes in the attribute. Unfortunately, this assumption is not justified. Summed questionnaire scores are not measures. At best, they provide ordinal rankings.

An ordinal scale allows observations to be placed in rank order, but it does not establish the magnitude of differences between observations. A patient with a score of twenty is ranked above a patient with a score of ten, but this does not imply that the first patient possesses twice as much of the attribute. Nor does it imply that the difference between ten and twenty is equivalent to the difference between forty and fifty. Ordinal scores support ordering, but they do not support arithmetic operations that require interval or ratio properties. Consequently, the widespread practice of summing questionnaire responses and treating the resulting score as a quantitative measure lacks a measurement foundation.

It was precisely this limitation that motivated the development of Rasch measurement theory. Rasch recognized that ordinal scores are not measures and that a different approach was required if latent attributes were to be assessed according to the standards of measurement applied elsewhere in science. Rather than assuming that responses could simply be added together, Rasch proposed a measurement model that evaluates whether observed responses conform to the requirements necessary for constructing a measure. The objective is not to create a score but to establish a lawful measurement scale.

The challenge is that latent attributes cannot be measured directly. They require a measurement instrument. This instrument is typically a questionnaire composed of carefully developed items intended to reflect different levels of the attribute. The purpose of the questionnaire is not to generate a score. Its purpose is to provide the observational evidence necessary for measurement. The responses themselves are not the measure. They are merely the raw material from which measurement may be established.

For a latent attribute to support claims regarding therapy impact, a number of conditions must be satisfied. The attribute must be clearly specified and unidimensional. The items must collectively define a single latent continuum. Response categories must function in an ordered manner. Item responses must exhibit local independence. The relationship between persons and items must be invariant across populations and circumstances. The resulting scale must permit meaningful comparisons among individuals and across time. Most importantly, the observations must satisfy the requirements of the Rasch measurement model.

Rasch analysis provides the framework through which these requirements can be evaluated. Unlike conventional questionnaire scoring, Rasch analysis does not assume that responses can be added together to create a measure. Instead, it tests whether the data satisfy the conditions necessary for measurement. Items that fail to contribute to the measurement of the attribute are removed or revised. Category functioning is evaluated to ensure that response options operate as intended. Differential item functioning is assessed to determine whether items perform consistently across relevant subgroups. Item fit statistics, local independence and residual analyses are examined to determine whether the requirements of measurement have been met.

The objective of this process is not to maximize statistical fit or predictive performance. The objective is to construct an instrument that satisfies the requirements necessary for lawful measurement. When these conditions are met, the resulting scale provides a Rasch logit ratio measure of the latent attribute. Individuals and items are located on a common continuum, allowing meaningful comparisons to be made regarding the extent to which an individual possesses the attribute.

A distinctive feature of Rasch measurement is that the outcome of interest is not a questionnaire score but the possession of a latent attribute. This distinction is critical. Conventional health outcomes research frequently evaluates therapy impact by comparing changes in average questionnaire scores before and after treatment. From a Rasch perspective, this is not the research question. The objective is not to determine whether a score has increased or decreased. The objective is to determine whether the extent to which individuals possess the latent attribute has changed.

Once a Rasch instrument has been calibrated, each individual can be located on a latent continuum defined by the attribute itself. For a pain instrument, the question is not whether the pain score has changed. Rather, it is whether the individual's location on the pain continuum has shifted. For a functioning instrument, the question is not whether a summed score has improved. Rather, it is whether the individual possesses a higher level of functioning than before treatment. The measure therefore refers directly to the attribute rather than to the numerical properties of a score.

At the population level, the research question is equally straightforward. The objective is to determine whether the distribution of possession of the latent attribute has shifted within the target population. A successful therapy is expected to move individuals to more favorable locations on the latent continuum. The central outcome is not score change but change in possession of the attribute. This may be expressed in terms of shifts in the population distribution, changes in average possession, or changes in the proportion of patients achieving specified levels of the attribute. In each case, the focus remains the same: possession of the latent trait rather than changes in questionnaire scores.

This perspective transforms the assessment of therapy impact. The Rasch instrument is not designed to generate numerical scores for statistical manipulation. It is designed to provide a lawful measure of a latent attribute that can support evaluable and falsifiable claims. The central question becomes: has therapy increased or decreased the extent to which patients possess the attribute of interest? Once this question is answered, the resulting claims can be tested, replicated and subjected to empirical scrutiny in exactly the same manner as claims based upon manifest attributes.

Importantly, the Rasch framework does not relax the standards required for measurement. It provides a means of satisfying them when the attribute cannot be observed directly. Just as manifest attributes require clearly specified, unidimensional ratio measures capable of supporting evaluable and falsifiable claims, latent attributes require clearly specified, unidimensional Rasch measures capable of supporting the same scientific objectives. The difference lies not in the standards but in the method through which measurement is achieved.

For students of HTA the implications are straightforward. The measurement of latent attributes cannot be reduced to questionnaire scoring, preference weighting or the summation of responses. Measurement requires the development of an instrument that satisfies the requirements of Rasch measurement theory and yields a Rasch logit ratio scale. Once such a measure has been established, claims regarding therapy impact can be evaluated, replicated and potentially falsified in exactly the same manner as claims based upon manifest attributes.

The significance of this conclusion is profound. Once the distinction between manifest and latent attributes is recognized, the measurement problem becomes remarkably simple. Manifest attributes require linear ratio measures. Latent attributes require Rasch logit ratio measures. Together, these two frameworks provide the foundation for scientific assessment of therapy impact and eliminate the need for the utility-based constructions that dominate contemporary health technology assessment.

A final observation is unavoidable. Despite the central importance of latent attributes in health technology assessment, there is little evidence that Rasch measurement theory is taught within contemporary HTA programs. The interrogation of the HTA knowledge bases over 230 HTA knowledge bases provides compelling evidence that knowledge and understanding of Rasch measurement are essentially absent. Across all programs, there is no indication that students are systematically introduced to the distinction between ordinal scores and measures, the concept of latent attribute possession, the requirements for unidimensional measurement, the construction of Rasch logit ratio scales, or the role of invariance and specific objectivity in measurement. At the same time, students are routinely exposed to utilities, QALYs, preference-based measures and simulation models that depend implicitly upon assumptions regarding measurement.

This finding is particularly important because Rasch measurement theory represents the only framework capable of transforming observations regarding a latent attribute into a lawful measure. Without an understanding of Rasch principles, students are left with the mistaken impression that summing questionnaire responses creates a measure and that changes in summed scores represent changes in the underlying attribute. The result is a curriculum in which latent attributes are discussed extensively but the only scientifically defensible approach to their measurement remains largely invisible.

The implications extend beyond student education. If faculty themselves are unfamiliar with Rasch measurement and the concept of attribute possession, then successive generations of students will continue to inherit a framework in which ordinal scores are mistaken for measures and numerical constructions are accepted as evidence of therapy impact. The interrogation findings suggest that this is precisely the situation that currently exists. The absence of Rasch measurement from HTA education is therefore not a minor curriculum gap. It represents the absence of the only measurement framework capable of supporting lawful quantitative claims for latent attributes.

EVALUATING CLAIMS FOR THERAPY IMPACT

The ultimate purpose of measurement is not to generate numbers. The purpose of measurement is to support claims. In healthcare, manufacturers, clinicians, researchers and policy makers routinely make claims regarding the impact of therapies. A treatment is claimed to improve survival, reduce hospital admissions, lessen pain, improve functioning, increase adherence or enhance quality of life. The role of health technology assessment is to determine whether such claims are true. This objective may appear straightforward, but it immediately raises a fundamental question: what standards must a claim satisfy before it can be regarded as scientifically credible?

The answer begins with a principle that has guided scientific inquiry for centuries. Claims must be capable of empirical evaluation. It is not sufficient to propose that a therapy produces a desirable outcome. The outcome must be defined, measured and subjected to observation. Without measurement there can be no evaluation. Without evaluation there can be no replication. Without replication there can be no falsification. Scientific assessment therefore begins not with statistical analysis or modelling but with the specification of a claim that can be tested against evidence.

For a claim regarding therapy impact to be evaluable, several conditions must be satisfied. First, the target population must be identified. Second, the therapy and comparator must be specified. Third, the attribute of interest must be clearly defined. Fourth, the measure employed must satisfy the requirements of measurement. Fifth, the time horizon over which the claim is to be evaluated must be stated. Finally, there must be a criterion by which the claim can be judged successful or unsuccessful. These requirements transform a general assertion into a proposition capable of empirical examination.

The importance of identifying the attribute cannot be overstated. Every claim concerns an attribute. A therapy may reduce hospital admissions, increase treatment persistence, improve survival, lessen fatigue or improve physical functioning. Each claim refers to a specific attribute. Unless that attribute is identified, there is no basis for measurement. Equally important, the attribute must be classified as either manifest or latent. This distinction determines the type of measurement required and the form of evidence that can be generated.

For manifest attributes, evaluation is relatively straightforward. Consider a claim that a therapy reduces hospital admissions by twenty percent over twelve months. Hospital admissions are directly observable. They can be counted and expressed on a ratio scale. The claim can therefore be tested by observing admissions within the target population and comparing outcomes between treatment groups. The resulting evidence is transparent, replicable and potentially falsifiable. If admissions do not decline, the claim fails. If admissions decline by less than expected, the claim may require modification. The measure provides a direct link between observation and inference.

The same logic applies to survival, medication possession, emergency department visits, physician consultations and numerous other manifest outcomes. Because these attributes are measured using linear ratio scales, arithmetic operations are admissible and claims can be expressed in terms of observable changes in magnitude. The assessment of therapy impact therefore becomes a matter of empirical investigation rather than speculation.

Latent attributes require a different measurement framework but the same scientific standards. Consider a claim that a therapy reduces fatigue or improves physical functioning. Neither attribute can be observed directly. They must be measured through a Rasch instrument designed to estimate the extent to which individuals possess the attribute. The purpose of the measure is not to generate a score. The purpose is to estimate possession of the latent trait.

This distinction has important implications for the assessment of therapy impact. Conventional outcomes research often focuses on score change. Patients complete a questionnaire before and after treatment and differences in average scores are interpreted as evidence of improvement. From a Rasch perspective, this is not the central research question. Summed questionnaire scores are ordinal rankings, not measures. They indicate order but not magnitude. Consequently, changes in scores cannot automatically be interpreted as changes in the attribute itself.

The Rasch framework adopts a different perspective. The outcome of interest is the possession of the latent attribute. Each individual is located on a latent continuum defined by the attribute being measured. The research question is therefore whether therapy changes the extent to which individuals possess that attribute. At the population level, the objective is to determine whether the distribution of possession shifts within the target population. A successful therapy should move patients to more favorable locations on the latent continuum, increasing average possession of desirable attributes or reducing possession of undesirable attributes.

This approach places latent and manifest attributes on a common scientific footing. For manifest attributes, the question is whether the magnitude of the attribute changes. For latent attributes, the question is whether the extent of possession changes. In both cases, the objective is to evaluate a claim against observation. The distinction lies not in the scientific standard but in the measurement framework employed.

The importance of evaluation leads naturally to the concept of falsification. A claim that cannot be shown to be false is not a scientific claim. Scientific progress depends upon the ability to compare predictions with observations and revise conclusions when evidence contradicts expectations. A therapy claim must therefore expose itself to the possibility of failure. This requirement is as important as measurement itself. The purpose of measurement is not merely to describe outcomes but to create the conditions under which claims can be challenged.

These principles have important implications for contemporary health technology assessment. Much of HTA is organized around utilities, quality-adjusted life years and simulation models. Yet these constructs begin not with attributes but with numerical representations of health-state preferences. The central question becomes the value assigned to a health-state description rather than the measurement of a manifest or latent attribute. Utilities are generated through preference elicitation exercises, transformed through econometric models and combined with time to create QALYs. At no stage is there a requirement to identify a unidimensional attribute or establish a lawful measure according to the standards described in the previous sections.

The consequence is that utilities, QALYs and simulation outputs occupy a different scientific status from claims based upon manifest or latent measures. A claim regarding hospital admissions can be evaluated directly. A claim regarding possession of a latent attribute can be evaluated through Rasch measurement. In contrast, a claim expressed as a gain in QALYs depends upon a sequence of assumptions regarding preferences, utility construction and arithmetic operations. The connection between the claim and the attribute of interest becomes increasingly obscure.

Once the distinction between manifest and latent attributes is recognized, the assessment of therapy impact becomes remarkably straightforward. Manifest attributes require linear ratio measures. Latent attributes require Rasch logit ratio measures. In both cases, claims can be expressed in a form that permits empirical evaluation, replication and falsification. The complexity associated with utilities, QALYs and simulation models becomes unnecessary because the focus returns to the attribute itself rather than to numerical constructions derived from preferences.

The future of health technology assessment therefore depends upon a return to first principles. The objective is not to produce increasingly sophisticated models or more elaborate numerical outputs. The objective is to evaluate claims regarding therapy impact. To accomplish this, claims must be linked to attributes, attributes must be measured according to the standards appropriate to their nature, and the resulting propositions must be exposed to empirical scrutiny. Only then can HTA satisfy the requirements of science and contribute to the continuing growth of objective knowledge.

CONCLUSION

If you are a faculty member or curriculum committee member responsible for a HTA program, the implications of this paper are straightforward. The purpose of pharmacy education is not merely to familiarize students with the methods currently employed in health technology assessment. It is to equip them with the knowledge necessary to evaluate those methods critically and to determine whether the claims they generate satisfy the standards of scientific inquiry.

The preceding sections have outlined what may be regarded as the minimum measurement competencies required for the assessment of therapy impact. Students should understand that measurement precedes arithmetic. They should understand Stevens' scales of measurement and the distinction between nominal, ordinal, interval and ratio scales. They should be familiar with the axioms of representational measurement and the requirement that quantitative claims be grounded in unidimensional attributes. They should appreciate the distinction between manifest and latent attributes and recognize that each requires a different measurement framework.

Students should also understand that manifest attributes require linear ratio measures characterized by a true zero, ratio comparisons and admissible arithmetic operations. Equally important, they should understand that latent attributes cannot be measured through the simple summation of questionnaire responses. Summed scores are ordinal rankings, not measures. The measurement of latent attributes requires a Rasch instrument capable of transforming observations into a Rasch logit ratio scale that supports meaningful quantitative comparisons and evaluable claims regarding the possession of the attribute.

Most importantly, students should understand that the purpose of measurement is to support evaluable and falsifiable claims. The objective of health technology assessment is not the generation of numerical outputs, utility scores or simulation results. The objective is the evaluation of claims regarding therapy impact. Those claims must be linked to clearly specified attributes, measured according to the standards appropriate to those attributes and subjected to empirical testing in real populations.

The interrogation of the 230 HTA knowledge bases detailed on the Maimon Research website suggests that these principles are largely absent from contemporary HTA education. Across programs there is little evidence that students are introduced to Rasch measurement theory, the concept of latent attribute possession, the distinction between ordinal scores and measures, or the conditions required for lawful quantitative claims. Yet these concepts are not advanced topics. They are the foundations upon which scientific assessment depends.

The challenge facing HTA education is therefore not simply whether utilities, QALYs or simulation models should continue to be taught. The more fundamental question is whether students are being provided with the measurement foundations necessary to evaluate those constructs critically. If the answer is no, then graduates are being asked to accept quantitative claims without understanding the standards those claims are required to satisfy.

The standards described in this paper should therefore be regarded as core elements of every HTA curriculum concerned with health outcomes research and health technology assessment. They are not optional additions to existing coursework. They are the foundation upon which the scientific evaluation of therapy impact must rest. Without them, students may become proficient in the use of numerical techniques while remaining unfamiliar with the principles that determine whether those techniques support meaningful measurement. With them, future practitioners, researchers and policy analysts will be equipped to evaluate claims according to the standards of evidence, replication and falsification that define scientific inquiry.

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.

REFERENCES

[1] Stevens S. On the Theory of Scales of Measurement. *Science*. 1946;103(2684):677-80

[2] Krantz D, Luce R, Suppes P, Tversky A. Foundations pf Measurement Vol 1: Additive and Polynomial Representations. New York: Academic Press, 1971

[3] Bond T, Zi Yan, Heene M. Applying the Rasch Model: Fundamental Measurement in the Human Sciences (4th Ed). New York: Routledge, 2021

[4] Rasch G, Probabilistic Models for some Intelligence and Attainment Tests. Chicago: University of Chicago Press, 1980 [An edited version of the original 1960 publication]

[5] Wright B. Solving measurement problems with the Rasch Model. *J Educational Measurement.* 1977;14(2):97-116