# MAIMON RESEARCH LLC

# ARTIFICIAL INTELLIGENCE LARGE LANGUAGE MODEL INTERROGATION

# THE MEASUREMENT FUTURE IN HEALTH TECHNOLGY ASSESSMENT

# UNITED STATES: 40 YEARS OF QALY MEASUREMENT FAILURE AND THE END OF THE LEGACY HTA CURRICULUM

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

**LOGIT WORKING PAPER No 5900 SEPTEMBER 2026**

**www.maimonresearch.com**

**Tucson AZ**

## ABSTRACT

*For more than forty years, health technology assessment (HTA), health economics and health economics and outcomes research (HEOR) in the United States have developed an increasingly sophisticated analytical framework incorporating health-state preferences, utilities, quality-adjusted life years (QALYs), incremental cost-effectiveness ratios (ICERs), reference-case modelling and simulation. This paper argues that this development proceeded without establishing the logically prior requirements of representational measurement. More than 40 U.S. Logit papers examining measurement failure and measurement inversion, together with a program-content assessment reporting 0/17 representational-measurement topics and 20/20 contemporary HTA and health-economics topics, provide evidence of a fundamental inversion: arithmetic and modelling have preceded demonstration of measurement.*

*The QALY provides the decisive test. The construction QALY = T × U requires the health-state preference value U to function as a proportional adjustment to ratio-measured time. A preference value of 0.5, however, does not establish that anything has been measured as one-half. Six and twelve months are commensurable ratio measures within Time Space, permitting the lawful proportion 6/12 = 0.5. A value of 0.5 originating in Preference Space does not thereby acquire the measurement meaning of that time ratio. Numerical identity is not measurement identity. Dimensional homogeneity reinforces the failure: if the QALY retains the dimension of time, U must be dimensionless, but dimensionless does not establish a lawful proportional measure. Even a hypothetical lawful preference ratio would remain a preference ratio and could not become a time ratio merely because the same numeral represented both. The QALY therefore cannot exist as a lawful measure. No lawful QALY means no lawful incremental QALY and no scientifically meaningful cost-per-QALY ICER.*

*The implications extend to education. Legacy HTA has been internationally reproduced as a methodological memeplex through universities, professional organizations, journals, guidelines and decision institutions. Replication of methodology, however, is not replication of its measurement foundation. Faculty have a duty to ensure that students are aware of this measurement challenge and equipped to evaluate it.*

*The required response is abandonment rather than repair. A nine-unit reconstruction begins with defined attributes, distinguishes manifest from latent measurement, establishes lawful measurement before admissible arithmetic, and requires therapy-impact claims to be empirically evaluable, replicable and falsifiable. Evidence gaps justify empirical inquiry, not imaginary numerical substitutes for evidence. Measurement must precede arithmetic if HTA is to return to the standards of normal science.*


## UNITED STATES: FORTY YEARS OF LEGACY HTA

United States health technology assessment (HTA), health-economics teaching, research and practice now face an argument that goes to the foundations of the discipline. Legacy HTA presents itself as a quantitative framework while failing to establish the measurement foundations required for its central quantitative claims: representational measurement and the scales of measurement [1] [2]. More than 40 U.S. logit papers have examined aspects of measurement failure, measurement inversion, curriculum inversion and the apparent absence of lawful measurement standards from the U.S. HTA and health-economics knowledge base [3] . Yet the established framework of health-state preferences, utilities, QALYs, cost-per-QALY claims and economic modelling remains deeply embedded in U.S. health economics, HEOR and graduate education.

If the United States is to make the necessary transition to an HTA curriculum grounded in lawful measurement, a catalyst may now be available. A recent evaluation of the measurement basis of the QALY demonstrates that the problem is not merely that the QALY is imperfect, controversial or inadequately validated. Once the standards of representational measurement are applied, the QALY is found never to have existed as a lawful measurement construct [4] More importantly, there was no missing measurement proof that might eventually have justified it. There was no possible proof. Given the nature of the preference scores employed and the quantitative function assigned to them, the procedure itself is inadmissible.

For more than 40 years, an apparent failure to recognize the requirements of representational measurement and the constraints imposed by scales of measurement allowed the QALY and associated cost-utility framework to become established components of U.S. health economics and HEOR. This framework has been reproduced through university teaching, research programs, professional education, economic evaluation, value assessment and decision modelling. The more than 40 U.S. Logit papers provide an accumulating body of evidence that the measurement assumptions underlying this inheritance were not subjected to the logically prior standards required for lawful quantitative claims.

This conclusion is reinforced by the U.S. program-content assessment [5]. The result is the same striking inversion identified in other national HTA environments: 0/17 for representational-measurement content and 20/20 for contemporary HTA and health-economics content. The U.S. knowledge base therefore presents a highly developed downstream analytical apparatus accompanied by an apparent absence of the measurement framework required to determine whether its central quantitative operations possess empirical legitimacy.

The contrast is particularly striking because the United States cannot plausibly be characterized as lacking methodological sophistication. Its universities, schools of pharmacy, schools of public health, medical schools, health-economics programs and research centers encompass economic evaluation, cost-effectiveness and cost-utility analysis, health-state preferences, QALYs, ICERs, patient-reported outcomes, comparative effectiveness research, decision analysis, Markov modelling, microsimulation, uncertainty analysis, budget impact, value assessment, real-world evidence and resource-allocation decision making. The analytical apparatus is extensive and technically sophisticated. What the program-content interrogation fails to identify is its logically prior measurement gatekeeper.

This is an extraordinary achievement in the history of the quantitative sciences. HTA developed an increasingly sophisticated analytical apparatus while largely ignoring the measurement requirements governing whether its numbers represented measures and whether the interpretations placed upon its arithmetic were empirically meaningful. These are not recently invented objections. The development of measurement theory and the formal analysis of scales established standards decades ago requiring the properties of numerical representations to be established before the quantitative interpretations placed upon them can be justified. Dimensional considerations impose additional constraints upon relationships between quantities.

The QALY now exposes the consequences with unusual clarity. The familiar construction: QALY = T × U requires the health-state preference value U to perform a proportional adjustment to time [6]. If U = 0.5 and T = 12 months, the construction produces six quality-adjusted months. But the logically prior measurement question is elementary: What has been measured as one-half? A preference score bounded between zero and one does not acquire proportional meaning merely because 0.5 appears numerically to represent a half.

The problem is deeper than the absence of ratio properties in conventional health-state preference scores. If six months and twelve months are lawful ratio measures of duration, then: 6 / 12 = 0.5 is a legitimate proportional relationship within Time Space. A value of 0.5 originating in Preference Space does not thereby become that time proportion. Equality of the numerals does not establish equality of measurement meaning. Numerical identity is not measurement identity.

Nor is there an escape through constructing a better preference instrument. Even if two preference values were lawful ratio measures of a defined preference attribute, their ratio would remain a proportion within Preference Space. It would not become a ratio of durations merely because the same numerical value could be produced in Time Space. Even a lawful preference ratio would remain a preference ratio. It cannot become a time ratio merely because both are represented by the same number.

The implications for the United States are therefore no longer confined to methodological debate. More than 40 U.S. logit papers have examined the broader measurement problem. The 0/17 versus 20/20 program-content finding demonstrates its educational expression. The demonstration that the QALY itself cannot exist as a lawful measure now exposes the consequence at the center of the legacy framework.

The cascade is unavoidable. If there is no lawful QALY, there can be no lawful incremental QALY. If the denominator has no defensible measurement status, division of incremental costs by that quantity cannot produce a scientifically meaningful cost-per-QALY ICER. QALY-based modelling cannot repair the failure because the failure occurs before modelling begins. Simulation, probabilistic sensitivity analysis and increasingly elaborate decision models cannot manufacture measurement properties absent from their inputs.

For the United States, this has particular significance because there is no single national HTA authority whose methodological conversion must precede educational change. Universities, pharmacy schools, health-economics programs and research centers do not need to wait for a national agency, professional organization or another university to move first. Compliance with the requirements of measurement does not require institutional permission. Each program can ask independently whether the quantitative constructs it teaches satisfy the measurement requirements necessary for the arithmetic and interpretations assigned to them.

The educational consequences are therefore direct. U.S. universities cannot resolve the problem by adding a lecture on measurement theory to an otherwise unchanged HTA or HEOR curriculum.

Measurement is not topic number 38. It is the gatekeeper that determines whether the other quantitative topics survive scientific scrutiny. Once measurement precedes arithmetic, the starting point must become the definition of the attribute, determination of whether it is manifest or latent, construction of the appropriate lawful measure, identification of admissible arithmetic and formulation of therapy-impact claims capable of empirical evaluation, replication and falsification.

This places faculty in an uncomfortable but unavoidable position. Most were educated within the same legacy framework they now teach. They cannot reasonably be criticized for failing to teach measurement standards that were absent from their own graduate education. That explains how the legacy was transmitted; it cannot justify transmitting it again. Once the measurement failure has been exposed, which it has, another generation of U.S. students cannot simply be taught utilities, QALYs and cost-per-QALY analysis as legitimate quantitative science as though the problem had never been identified.

The more than 40 U.S. logit papers and the program-content assessment therefore establish a sequence of evidence that is becoming increasingly difficult to dismiss. The first stage identified measurement failure and measurement inversion. The second exposed their reproduction through the U.S. knowledge base and program content. The QALY analysis now demonstrates the consequence of allowing that inversion to persist: a foundational quantity employed for decades that could never satisfy the measurement requirements necessary for the quantitative function assigned to it.

The question facing U.S. universities and research centers has consequently changed. It is no longer whether representational measurement deserves greater attention within the established HTA curriculum. Nor is the appropriate response to preserve the existing curriculum while adding measurement as another methodological topic. The issue is whether the legacy framework can survive once representational measurement is restored to its logically prior position.

A practical alternative is available now. The nine-unit Maimon Research program reconstructs HTA from measurement forward [7]. It begins with defined attributes and distinguishes the measurement requirements of manifest and latent attributes. Manifest attributes require appropriate linear ratio measurement; latent attributes require a measurement model capable of constructing the required unidimensional measurement structure. Measurement properties determine admissible arithmetic, and quantitative therapy-impact claims must be capable of empirical evaluation, replication and falsification. The sequence is reversed from that of legacy HTA: attribute → measurement → scale → admissible arithmetic → quantitative claim.

This transition to lawful measurement is not an option. It is the only scientifically defensible response to the failure of legacy HTA. It provides students with the measurement foundations required to distinguish the standards of ratio measurement for manifest attributes from the fundamentally different requirements for measuring latent attributes. Measurement once again becomes the gatekeeper for arithmetic and quantitative claims. For the first time in forty years, HTA can return to the standards of normal science.

The implications are uncompromising. The entire edifice of legacy HTA including health-state valuation, utilities, QALYs, cost-per-QALY claims and reference-case simulation modelling fails to meet the required standards of lawful measurement**.** Students will certainly want to know why constructs they may have been taught as foundations of quantitative HTA cannot survive those standards. A curriculum founded instead on defined attributes, lawful measurement and admissible arithmetic provides precisely the framework required to answer that question. There is nothing after forty years of legacy HTA that requires preservation merely because it has become established practice. Its foundational quantitative framework has no demonstrated scientific merit and cannot support the requirements of normal science: empirically evaluable claims, replication, falsification and the progressive evolution of objective knowledge concerning therapy impact**.** The task facing U.S. HTA education is therefore not to repair the legacy framework. It is to abandon it and reconstruct HTA from lawful measurement.

## UNDERSTANDING MEASUREMENT

The distinction between numbers and measures is fundamental. Any collection of numbers can be added, subtracted, multiplied or divided computationally. The ability of a calculator or computer to perform an arithmetic operation does not establish that the result has an empirical interpretation. Numbers become measures only when they represent an attribute through a measurement structure possessing defined properties. Those properties determine which transformations, comparisons and arithmetic interpretations are admissible.

Ratio measurement is critical whenever proportional meaning is claimed. A lawful ratio measure possesses equal units and a non-arbitrary zero, with admissible similarity transformations of the form $x' = ax$, $a > 0$. Statements such as one quantity being twice another or one-half another depend upon that ratio structure. A number does not acquire proportional meaning merely because it has been assigned a zero, bounded between zero and one, or described as a utility.

This distinction is fatal to the conventional QALY. Time can be represented on a ratio scale. Twelve months is twice six months, and the proportional relationship is preserved under admissible changes of the time unit. A health-state preference score does not acquire corresponding ratio properties merely because it is expressed numerically. In particular, assigning a composite health-state valuation the number 0.5 does not establish that one-half of a defined attribute has been measured.

Legacy HTA and health economics nevertheless proceed as though this distinction did not matter. Preference scores are treated as utilities; utilities are multiplied by time; QALYs are subtracted to create incremental QALYs; and incremental costs are divided by these quantities to create cost-per-QALY ICERs. These operations are incorporated into economic evaluations, value assessments, decision models, formulary submissions and other analyses intended to inform coverage, reimbursement and resource-allocation decisions. The arithmetic becomes progressively more elaborate while the measurement failure at its foundation remains untouched.

The issue is not whether analysts can perform the accepted calculations; they undoubtedly can. The prior question is whether the numbers entering those calculations are measures possessing the properties required for the quantitative claims being made. A correctly executed cost-utility model cannot establish retrospectively that its utility input possesses the measurement properties required for its assigned quantitative role. Probabilistic sensitivity analysis cannot manufacture a lawful

QALY. Increasing computational sophistication cannot repair a measurement failure that occurred before the arithmetic began.

The QALY exposes an even deeper problem. If six months and twelve months are lawful ratio measures of time, then 6 / 12 = 0.5 is a lawful proportional statement within Time Space. The numerator and denominator are commensurable measures of the same attribute. By contrast, a health-state preference value of 0.5 originates in Preference Space. Its numerical equality with the time ratio 6/12 does not give it the measurement meaning of that time ratio. Numerical identity is not measurement identity.

This problem cannot be escaped by proposing a better preference instrument. Even if preference values could be established as lawful ratio measures of a defined preference attribute, a ratio between those values would remain a proportion within Preference Space. It would not thereby become the proportional relationship required within Time Space.

Even a lawful preference ratio would remain a preference ratio. It cannot become a time ratio merely because both are represented by the same number.

This is measurement inversion: arithmetic is performed first and the measurement legitimacy of the quantities upon which it depends is either assumed or left unexamined. The problem is therefore not an isolated methodological oversight. It is embedded in the knowledge and educational framework through which legacy HTA and health economics have been taught, researched and reproduced.

The consequences extend well beyond university teaching. A quantity that fails the measurement test in a university classroom does not acquire measurement legitimacy when incorporated into a formulary submission, an economic model, a value-assessment framework, a methodological guideline or a reimbursement decision. Institutional acceptance cannot confer measurement properties upon a quantity that does not possess them.

The educational implications are particularly important. Generations of students have been taught increasingly sophisticated methods for manipulating utilities, QALYs, ICERs and modelled outcomes without first being required to establish whether the quantities involved possess the measurement properties necessary for those operations. Faculty may reasonably respond that they are teaching the framework they themselves were taught. That explains how the legacy was transmitted; it cannot justify transmitting it again. Once the measurement failure has been exposed, another generation of students cannot simply be taught the same framework as though the problem had never been identified.

Nor can the problem be resolved by inserting measurement theory as another topic within an otherwise unchanged curriculum. Measurement is not simply one methodological subject among economic evaluation, utilities, QALYs, ICERs, decision modelling and uncertainty analysis. It is logically prior to them. It is the gatekeeper for everything that follows. The sequence must begin with the definition of the attribute, identification of whether it is manifest or latent, establishment of the appropriate measurement structure and scale properties, determination of admissible arithmetic, and only then construction of quantitative therapy-impact claims capable of empirical evaluation, replication and falsification.

Measurement must precede arithmetic. Once that requirement is imposed, the conventional QALY becomes a critical test that legacy HTA cannot pass. The issue is not the sophistication of the health-economics and HTA infrastructure within which these methods are taught and applied. The issue is whether that infrastructure can demonstrate the measurement legitimacy of the quantitative constructs upon which its central claims depend. Once the measurement failure has been exposed, the question for universities, research centers and HTA programs is what they choose to teach next.

## THE FAILURE EXTENDS BEYOND THE QALY

A further reason for abandoning legacy HTA is that the failure is not confined to the QALY. The entire analytical framework proceeds without ratio measurement being established at the stages where proportional interpretation and arithmetic require it. The problem begins with the valuation of composite health states. Procedures such as the time trade-off and standard gamble generate numerical preference values, but the assignment of numbers does not establish a unidimensional attribute, equal units, a non-arbitrary zero or invariant ratios. Valuation is treated as though it were measurement.

It is almost as though the scales of measurement had never been considered. Stevens made the governing distinction explicit in his 1946 paper *On the Theory of Scales of Measurement*. Nominal, ordinal, interval and ratio scales do not differ merely in the kinds of numbers attached to observations; they differ in the empirical structure represented, the transformations that preserve that structure and, consequently, the quantitative interpretations that are meaningful. Ratio statements and the multiplicative and divisive interpretations associated with them require ratio-scale structure. A ratio scale requires an invariant zero and is preserved under similarity transformations of the form $x' = ax$, $a > 0$. This was not an obscure qualification discovered after the development of HTA. The scale distinction was part of the established measurement literature decades before the QALY was introduced. Yet legacy HTA proceeded to multiply preference values by time and divide costs by constructed outcomes without first demonstrating that the quantities being assigned these proportional roles possessed the required measurement properties. The measurement question that should have preceded the arithmetic was effectively bypassed.

The failure to appreciate the constraints imposed by ratio measurement was then repeated at virtually every stage of the legacy HTA analytical process. Multiattribute health-state frameworks combined responses across distinct dimensions and supplied these numerical composites as inputs to subsequent statistical analyses without first establishing a unidimensional measured attribute possessing the required scale properties. Regression models were fitted to multiattribute health-state values and used to construct utility algorithms, but statistical estimation cannot manufacture ratio measurement from inputs that lack ratio structure. The resulting algorithms were manipulated, transformed and applied as though their outputs possessed proportional meaning, while the requirements of a non-arbitrary zero, equal units and invariant ratios remained unaddressed. Preference elicitation through time trade-off and standard gamble likewise generated numerical valuations that were subsequently treated as utilities without demonstrating the ratio properties required for their assigned quantitative role.

Finally, these utilities, QALYs and other constructed numerical inputs were passed into decision trees, Markov models, microsimulations and probabilistic simulation models with no prior demonstration that they satisfied the ratio-scale requirements demanded by the arithmetic performed upon them. The same measurement failure was therefore reproduced sequentially: from multiattribute description, through statistical estimation and preference valuation, to utility algorithms, QALYs, ICERs and simulation. At every stage, increasingly complex arithmetic was permitted to substitute for the logically prior demonstration of measurement.

The problem is then compounded as each inadequately founded quantity becomes an input to the next stage of analysis. Preference values are redesignated as utilities and assigned a new quantitative function as proportional adjustments to ratio-measured time. Utilities are multiplied by duration to manufacture QALYs; QALYs are subtracted to produce incremental QALYs; and incremental costs are divided by these quantities to generate cost-per-QALY ICERs. These outputs are then embedded in decision trees, Markov models, microsimulations and probabilistic simulations. The computational sophistication increases at every stage, but the original measurement question remains unanswered. The downstream complexity does not strengthen the measurement foundation; it merely carries the original failure further into the analytical structure.

At no subsequent stage can these failures be repaired retrospectively. A simulation model cannot confer ratio properties upon an input that never possessed them. Probabilistic sensitivity analysis can characterize uncertainty around assigned numerical values, but it cannot establish that those numbers are measures. Monte Carlo simulation can repeatedly sample from specified distributions, but repetition cannot manufacture measurement properties. A Markov model can propagate numerical inputs through thousands of transitions without establishing that the quantities being accumulated, multiplied or compared possess the scale properties required for those operations. Computational excellence and measurement legitimacy are independent properties.

This has an important implication for the conventional defense of legacy HTA. The problem cannot be solved by improving the utility instrument, estimating a different tariff, collecting a larger preference sample, introducing a more sophisticated regression model or constructing a more elaborate simulation. Each of these responses operates downstream of the logically prior question: what attribute has been measured, on what scale, and what arithmetic does that scale permit? If that measurement gate has not been passed, subsequent analytical sophistication is irrelevant to the legitimacy of the quantitative claim.

The QALY is therefore not an isolated defect within an otherwise scientifically defensible analytical framework. It is the most conspicuous product of a measurement inversion that runs through the entire legacy HTA analytical sequence**.** The problem begins before the QALY is constructed and continues after it has been incorporated into cost-effectiveness models. From multiattribute health-state description, through preference elicitation, statistical estimation and utility algorithms, to QALYs, incremental QALYs, ICERs and simulation models, the properties required for measurement have been assumed, ignored or bypassed while arithmetic has proceeded.

For more than four decades, tens of thousands of teachers, researchers and analysts have accepted the proposition that the QALY can simply be defined as the product of time and a health-state preference value: QALY = T × U.

But definition cannot create measurement properties. The preference score (U) originates as a valuation attached to a composite, multiattribute health state. It is then assigned an entirely different quantitative role: it is treated as though it were a lawful proportional adjustment to ratio-measured time. Thus, if U = 0.5, twelve months is declared equivalent to six "quality-adjusted" months.

The obvious question is: what has been measured as one-half? The answer is nothing. A composite preference valuation bounded between zero and one does not thereby become a ratio measure. It has not acquired a non-arbitrary zero, equal units or invariant ratios. The number 0.5 therefore cannot mean one-half in the measurement sense required for a proportional adjustment to time.

The problem is deeper than inadequate validation. Neither representational measurement theory nor the dimensional requirements governing quantitative relationships can justify taking a preference-based number and treating it as a proportional adjustment to ratio-measured time. The preference valuation and elapsed time are fundamentally different quantitative objects. Calling the former "utility," declaring it dimensionless, or bounding it between zero and one cannot confer the measurement properties required for the quantitative role assigned to it.

This is the sleight of hand at the center of the QALY. A valuation of a composite health state is transformed by definition into something it has never been demonstrated to be and, given its construction, cannot be: a lawful proportional measure capable of adjusting time. Twelve months multiplied by 0.5 certainly produces the number six. What has not been produced is a measured quantity corresponding to six months of "quality-adjusted" time.

The failure becomes still clearer when the two quantitative spaces are distinguished. Six months and twelve months can be lawful ratio measures of the same attribute, duration. Their ratio: 6/ 12 = 0.5 \is therefore a lawful proportional statement within Time Space. By contrast, a health-state preference value of 0.5 originates in Preference Space. Numerical equality between these two values does not establish equality of measurement meaning. Numerical identity is not measurement identity.

Nor can the QALY be rescued by constructing a better preference instrument. Even if preference values could be established as lawful ratio measures of a defined preference attribute, their ratio would remain a proportion within Preference Space. It would not thereby become the proportional relationship required within Time Space.

Even a lawful preference ratio would remain a preference ratio. It cannot become a time ratio merely because both are represented by the same number. There is consequently no mystery awaiting resolution by additional empirical research. There is no missing validation study, alternative tariff, larger preference sample or more sophisticated statistical model that can supply the measurement legitimacy absent from the construction itself. There was no missing proof. There was no possible proof.

The QALY has therefore never existed as a lawful measurement construct. What has existed is a numerical convention in which a preference-based valuation is multiplied by elapsed time and the product is given the name “quality-adjusted life years.” For more than forty years, this convention has been taught and applied throughout health economics and HTA and incorporated into economic evaluation, value assessment, modelling and decision-making as though its measurement status had already been established.

The implications are not confined to one HTA agency, one country or one methodological guideline. They reach across universities, health-economics and HTA programs, research centers, value-assessment organizations, pharmaceutical manufacturers, health systems, payers and formulary decision-making. A QALY that fails the measurement test in a university classroom does not acquire measurement legitimacy when it enters an economic model, value assessment, formulary submission or reimbursement decision.

The question facing universities and research centers is therefore no longer how the QALY might be improved. It is why another generation of students should continue to be taught a quantitative construction that cannot satisfy the measurement requirements necessary for its claimed interpretation. Faculty may reasonably respond that this is the framework they themselves were taught. That explains how the legacy was transmitted; it cannot justify transmitting it again.

Once the standards of representational measurement are applied, the choice becomes straightforward. If the QALY passes the measurement gate, there is nothing to abandon. If it fails that gate, there is nothing to rescue. The task for HTA and health-economics education is not to search for another way to validate the QALY. It is to abandon the measurement inversion that made the QALY possible and reconstruct therapy-impact assessment from lawful measurement.

## DIMENSIONAL HOMOGENEITY

Dimensional homogeneity provides a second and independent constraint upon quantitative operations. It is not sufficient that numbers can be multiplied or divided computationally. The quantities represented by those numbers carry dimensions, and the dimensional meaning of the resulting quantity must be preserved throughout the operation. A quantitative equation cannot acquire empirical meaning merely because its numerical calculation can be performed.

The Theory of Dimensions addresses the relationships between physical or quantitative attributes and the dimensions in which they are expressed. Quantities such as time, length, mass and velocity are not merely numbers; they possess dimensional identities that constrain the mathematical relationships in which they can meaningfully participate. Dimensionality is therefore a fundamental safeguard against meaningless arithmetic. An equation may be computationally executable while remaining scientifically inadmissible if the dimensions of its quantities are incompatible with the relationship being asserted. Multiplication and division create predictable dimensional relationships, while addition and subtraction require commensurable quantities. These constraints are independent of statistical sophistication or modelling technique. Before accepting any quantitative relationship, it is therefore necessary to ask not only whether its components are measures, but whether their dimensions permit the proposed relationship.

Two lawful ratio measures representing different attributes can be multiplied or divided. Distance divided by time produces velocity; mass divided by volume produces density. In such cases, the resulting quantity has a derived dimension determined by the dimensions of the quantities entering the operation. If distance has dimension [L] and time has dimension [T], then: Velocity = Distance / Time and [V] = [L] / [T] in dimensional theory terms. There is therefore no general requirement that the numerator and denominator represent the same attribute when the purpose is to construct a dimensioned derived quantity.

The requirement is different when division is claimed to produce a dimensionless proportion. If T1 and T0 are two lawful measures of time, then T1 / T0 has the dimensional representation [T] / [T] = 1. The time dimensions cancel. If T1 = 6 months and T0 = 12 months then T1 / T0 = 6 / 12 = 0.5. The resulting value of 0.5 has a clear empirical interpretation: the first duration is one-half of the second. Both quantities are lawful ratio measures of the same attribute. The resulting dimensionless number is therefore a lawful proportional measure.

In other words, if a dimensionless proportion is required, the numerator and denominator of the ratio must have the same dimension. Their dimensions then cancel. For an adjustment to time, the clearest lawful proportion is therefore constructed within Time Space itself. If (T_1) and (T_0) are ratio measures of duration, then (T_1/T_0) has dimension (T/T=1) and is dimensionless. Thus, (6/12=0.5) legitimately means that six months is one-half of twelve months. The value 0.5 has proportional meaning because it was created from two commensurable ratio measures of the same attribute. Dimensionless is therefore the consequence of a lawful ratio; it does not itself establish one**.** A number originating in another attribute space cannot acquire the meaning of a time proportion merely because it is also dimensionless or happens to have the same numerical value.

This distinction is critical. A dimensionless number is not necessarily a dimensionless proportional measure. Merely assigning a number between zero and one does not establish that it represents the ratio of two commensurable ratio measures. Nor does calling such a number a "utility" establish the proportional meaning required for it to function as an adjustment factor.

This is precisely where the QALY fails. If the QALY is claimed to retain the dimension of time, then: [Q] = [T]. Given Q = T × U dimensional homogeneity requires [Q] **=** [T] **×** [U]. Therefore, if [Q] **=** [T] then [U] **=** 1**.**

This establishes that U must be dimensionless. It does not establish that U is the dimensionless proportional adjustment factor required for time**.** For that interpretation, the proportion must be established within Time Space from commensurable ratio measures of duration. If $T_1$ is to represent an adjusted duration relative to $T_0$, the required proportional factor is $U = T_1 / T_0$ so that $T_1 = T_0 \times U$. Because both $T_1$ and $T_0$ are lawful ratio measures of time, their dimensions cancel: [T] **/** [T] **= 1** producing a dimensionless factor whose proportional meaning is nevertheless derived entirely from Time Space.

The conventional QALY does something fundamentally different. In Q = T × U T originates in Time Space while U originates in Preference Space. Demonstrating that U is dimensionless does not transform a preference-space value into a time-space proportion. If U = 0.5 is to mean that twelve months becomes six months, the required empirical relationship is $T_1 / T_0 = 6 / 12 = 0.5$.

A preference value that also happens to equal 0.5 does not establish that relationship. The dimensions may be formally compatible while the required measurement meaning is absent. Numerical identity is not measurement identity.

The conventional health-state preference score provides no such basis. It is a numerical valuation assigned to a composite, multiattribute health state. It is not produced as the ratio of two commensurable ratio measures of a defined attribute. Its dimensions have not cancelled to produce a lawful proportion. Bounding the score between zero and one merely produces a number within a numerical interval; it does not establish the measurement structure required for the statement "one-half."

The QALY therefore commits a fundamental substitution. It substitutes an ersatz dimensionless preference number for a dimensionless proportional measure and then treats the two as equivalent. They are not. The distinction can be stated simply: Dimensionless number ≠ dimensionless proportional measure ≠ lawful proportional adjustment to time.

This is why $U = 0.5$ cannot simply be applied to twelve months to claim six measured "quality-adjusted" months. The arithmetic $12 \times 0.5 = 6$ is computationally correct. But computational correctness does not establish measurement legitimacy. The QALY requires the preference score to perform a measurement function that its construction cannot support.

Dimensional homogeneity therefore does not rescue the QALY. It exposes the problem. If the product is claimed to remain a measure in time, the multiplier must possess the legitimate dimensionless proportional meaning required to adjust time. The conventional preference-based utility does not. Once again, the conclusion is unavoidable: The ability to multiply T by U demonstrates only that two numbers can be multiplied. It does not demonstrate that the resulting number is a measure of quality-adjusted time.

There are consequently two distinct failures. First, the preference score lacks the ratio measurement structure required to establish 0.5 as a lawful proportion. Second, the numerical value is nevertheless used as though it were the proportional adjustment required within the entirely different attribute space of time.

Representational measurement provides no authority for this procedure. Nor can dimensional notation rescue it. The fact that a preference score is treated as dimensionless does not establish that it is a lawful dimensionless proportion, still less that it possesses the proportional meaning required to adjust a ratio measure of duration.

The error can be expressed very simply: Preference number = 0.5 does not establish Time proportion = 0.5 and therefore cannot establish Tadjusted / T = 0.5. The QALY simply assumes incorrectly the equivalence required for its construction.

This is the fundamental category error at the heart of quality adjustment. A number created in Preference Space is assigned, without lawful measurement justification, the proportional meaning required in Time Space. The arithmetic can be performed; the measurement claim cannot.

## COMMENSURABILITY

The concept of commensurability, the quality of being measurable or comparable, makes the failure still clearer. A ratio expressing a proportion such as “one-half” obtains its empirical meaning from comparison of commensurable measures of the same attribute. If T1 and T2 are lawful ratio measures of duration, then: T1 / T2 = 0.5means that T1 is one-half of T2 in Time Space. The numerator and denominator are commensurable measures of time.

Similarly, even if it were assumed hypothetically that P1 and P2 were lawful ratio measures of the same preference attribute P1 / P2 = 0.5 would establish only that P1 is one-half of P2 in Preference Space. P1 and P2 would be commensurable with each other, just as T1 and T2 are commensurable with each other. But preference and time are not commensurable attributes. Consequently P1 / P2 = 0.5 does not establish:T1 / T2 = 0.5

The appearance of the same number on both sides cannot bridge the distinction between the attributes. The first 0.5 represents a proportional relationship between two preference measures; the second represents a proportional relationship between two time measures.

Numerical equality does not establish commensurability, and it does not establish measurement equivalence. This is precisely the operation required by the QALY. A number originating in Preference Space is treated as though it were the proportional adjustment required in Time Space. Even granting lawful ratio measurement of preference for the sake of argument cannot justify that substitution.

The problem is therefore not simply that conventional utility scores fail ratio measurement. Even a lawful preference ratio would remain a preference ratio. It cannot become a time ratio merely because both are represented by the same number.

## THE QALY CANNOT EXIST AS A LAWFUL MEASURE

The failure of the QALY is now unavoidable. It cannot be rescued by better preference instruments, alternative elicitation procedures, larger samples, different tariffs or more sophisticated statistical transformations. The problem is not that the conventional health-state preference score has failed to demonstrate adequate ratio properties. Even if, purely for argument, a lawful ratio measure could be constructed in Preference Space, the QALY would still fail. Numerical equality is not measurement equivalence.

This eliminates the final possible defense of the QALY. Giving preference values lawful ratio properties would not solve the problem. Ratio measurement establishes proportional meaning within the attribute space being measured. It does not authorize that proportion to be reassigned as the proportional adjustment of a different attribute. The QALY nevertheless requires exactly this unsupported substitution. Given QALY = T × U a value U = 0.5 is used to assert Tadjusted = 0.5 × T and therefore Tadjusted / T = 0.5. But the proportional relationship required here is a relationship in Time Space. A ratio established in Preference Space cannot, merely by numerical substitution, become a ratio in Time Space.

The conventional QALY is in an even worse position because its health-state preference scores are not lawful ratio measures of a defined unidimensional attribute in the first place. The construction therefore fails twice. It fails to establish the required proportional measurement properties for the preference number, and even if those properties were granted hypothetically, they would remain properties of a relationship in Preference Space rather than the required proportional relationship in Time Space.

There is therefore no escape route through better measurement of preference. The problem is inherent in the QALY construction itself. This has consequences far beyond the QALY. The QALY is the gateway quantity upon which the downstream architecture of legacy cost-utility HTA depends. If the QALY cannot exist as a lawful measure, neither can subtraction of QALYs manufacture a lawful incremental QALY: ΔQALY = QALY2 − QALY1. If the denominator is not a lawful measure of incremental therapy impact, subsequent division cannot manufacture a scientifically meaningful cost-per-QALY ICER ICER = ΔCost / ΔQALY.

Nor can a reference case repair the failure by specifying discount rates, model structures, time horizons, sensitivity analyses, uncertainty distributions or decision thresholds. These operations occur downstream of the failed QALY construction. Increasing their sophistication cannot create the measurement quantity that was never established upstream. The sequence is therefore inexorable: No lawful proportional adjustment of time → no lawful QALY → no lawful ΔQALY → no scientifically meaningful cost-per-QALY ICER → collapse of the QALY-based reference-case framework.

This is why the QALY problem cannot be accommodated as another methodological limitation of legacy HTA. There is nothing to validate, refine or repair. The attempted construction fails before the QALY comes into existence as a measure. Everything in legacy cost-utility HTA that requires the QALY as an input consequently loses its measurement foundation.

For more than forty years, legacy HTA has behaved as though assigning the same numeral to relationships in different attribute spaces makes those relationships interchangeable. It does not P1 / P2 = 0.5 does not establish T1 / T2 = 0.5. No amount of convention, institutional acceptance or computational sophistication can make that implication true. The QALY cannot exist as a lawful measure. Its attempted construction is abortive, and the legacy HTA framework built downstream upon it collapses with it.

## FAILED CURRICULUM CONTENT

The QALY measurement failure identified in the preceding sections cannot be separated from the way legacy HTA and health economics have been taught. The problem is not simply that an inadmissible quantitative construct entered economic evaluation. It is that generations of students could be taught health-state preferences, utilities, QALYs, ICERs and simulation modelling without first being given the representational-measurement framework required to determine whether those constructs could exist as lawful measures. The result is curriculum inversion: downstream analytical methods are taught before the measurement foundations required to establish their quantitative legitimacy.

A lawful curriculum would reverse this sequence. Before students manipulate numerical values, they should understand that measurement precedes arithmetic; that an attribute must first be defined; that scales of measurement determine admissible transformations and quantitative interpretations; that proportional claims require appropriate ratio structure; that dimensional homogeneity constrains quantitative relationships; and that manifest and latent attributes require fundamentally different measurement pathways. For latent attributes, students must understand the requirements for unidimensional measurement, invariance, person-item conjoint measurement and specific objectivity. These are not supplementary methodological topics. They determine whether subsequent arithmetic has empirical meaning.

The contrast with conventional HTA education is striking. Students may encounter economic evaluation, cost-effectiveness analysis, cost-utility analysis, pharmacoeconomics, utilities, health-state preferences, QALYs, ICERs, patient-reported outcomes, decision trees, Markov models, simulation, uncertainty analysis, budget-impact analysis, value assessment and reimbursement decision making. The analytical apparatus may be extensive and technically sophisticated. Yet sophistication downstream cannot compensate for the absence of the logically prior measurement gatekeeper upstream.

This creates an extraordinary educational inversion. Students can be introduced to utilities before establishing what constitutes a lawful measure; to QALYs before determining whether the preference multiplier possesses the proportional properties required of it; to ICERs before establishing whether their outcome denominator represents a legitimate quantity; and to increasingly sophisticated simulation models before being taught that arithmetic operations acquire empirical meaning only through the properties of the measures upon which they operate.

The responsibility to students extends beyond curriculum design. Faculty have a duty of care not to keep students unaware of a fundamental scientific challenge to the quantitative framework they are being taught. Faculty may disagree with the critique of legacy HTA and may choose to continue teaching utilities, QALYs, cost-per-QALY ICERs and reference-case modelling. What is no longer defensible is to teach these constructs as though their measurement legitimacy were uncontested. Students should be informed that a representational-measurement critique has been advanced which concludes that the central quantitative constructs of legacy HTA fail the requirements for lawful measurement and that, in the case of the QALY, the preference value cannot perform the proportional adjustment to time required by its construction.

This obligation does not require faculty to endorse the critique. It requires them to disclose it, explain its scientific basis and allow students to examine the arguments for themselves. If faculty believe legacy HTA survives the measurement challenge, they should demonstrate how: identify the attributes being measured, establish the required scale properties, justify the arithmetic performed and show that the resulting claims satisfy the requirements of normal science. Students can then evaluate the competing claims against explicit scientific standards.

What faculty were taught may explain why the measurement challenge was absent from earlier curricula. It cannot justify keeping another generation of students unaware of it once the challenge has been made explicit. Doubling down on legacy HTA while withholding the existence and substance of the measurement critique would deny students the intellectual opportunity to determine for themselves whether the methods they are being taught can withstand scientific scrutiny. The obligation is therefore straightforward: teach the measurement challenge, confront it openly and require legacy HTA to demonstrate its legitimacy. Students should accept no less.

The survival of the QALY becomes much easier to understand in such an educational environment. If students and faculty are not introduced to representational measurement, ratio measurement, dimensional homogeneity, commensurability and attribute-specific measurement requirements, they lack the conceptual framework necessary to identify the foundational error in the QALY construction. The QALY survived not because its measurement legitimacy had been demonstrated, but because the logically prior measurement questions required to challenge that legitimacy were largely bypassed.

The measurement failure is elementary. A preference value generated in Preference Space cannot simply be treated as the proportional adjustment required in Time Space. A value of 0.5 attached to a composite health-state valuation does not establish that anything has been measured as one-half. If six months and twelve months are lawful ratio measures of time, then $6 / 12 = 0.5$ establishes a lawful proportion between commensurable measures within Time Space. A preference value represented by the same numeral does not acquire that measurement meaning. Numerical identity is not measurement identity.

Even the hypothetical construction of a lawful preference ratio would not solve the problem. A preference ratio remains a preference ratio; it does not become a time ratio merely because both happen to be represented by the same number. There was therefore no missing measurement proof that legacy HTA had simply neglected to provide. Given the quantitative role assigned to the health-state preference value in $QALY = T \times U$ there was no possible proof capable of transforming that preference-space quantity into the required time-space proportion.

The consequences propagate throughout the curriculum. If the QALY cannot exist as a lawful measure, there is no lawful incremental QALY. Without a lawful incremental QALY, there is no scientifically meaningful cost-per-QALY ICER. Reference-case simulation, discounting, uncertainty analysis and additional modelling cannot restore a measurement foundation that failed before the QALY was created. A curriculum presenting these constructs as legitimate quantitative measures and outcomes cannot be repaired simply by increasing its methodological sophistication.

The curricular problem extends beyond the QALY itself. Multiattribute health-state frameworks generate numerical composites without establishing a single unidimensional measured attribute possessing the required ratio structure. Regression models and utility algorithms manipulate these values without creating the measurement properties absent from their inputs. Preference elicitation procedures generate valuations that are assigned proportional roles without demonstrating ratio measurement. These quantities are subsequently entered into simulation models whose mathematical sophistication can conceal, but cannot repair, the measurement failures inherited from earlier stages. The failure to impose the measurement gate is repeated throughout the analytical sequence.

This helps explain why curriculum failure becomes self-replicating. Faculty teach the framework they themselves were taught. Students inherit the same analytical structure. Those students become researchers, analysts and faculty and reproduce it for another generation. Methodological guidance, journals, professional organizations and institutional practice reinforce the same framework. Repetition then creates the appearance of validation when what has actually been replicated is the methodology rather than the measurement proof upon which its legitimacy should depend.

This process does not require incompetence or bad faith on the part of individual faculty. If representational measurement, ratio measurement, dimensional homogeneity, commensurability and the distinction between manifest and latent attributes were absent from their own education, they were denied the intellectual tools required to interrogate the measurement legitimacy of the framework they subsequently inherited and taught.

But that explanation has an important limit. What faculty were taught explains the past; it cannot determine what they teach next. Once the missing measurement framework has been identified and its consequences made explicit, inheritance can no longer justify reproduction. Another generation of students cannot be taught utilities, QALYs, ICERs and reference-case simulation as legitimate quantitative science without first being provided with the measurement standards required to determine whether those constructs and operations survive scientific scrutiny.

Nor can the problem be resolved by adding representational measurement as another topic within an otherwise unchanged curriculum. Measurement is not simply one methodological subject among economic evaluation, utilities, QALYs, ICERs, decision modelling and uncertainty analysis. It is logically prior to them. Representational measurement determines whether the quantities entering those downstream methods are measures, what transformations preserve their meaning, what arithmetic is admissible and what quantitative interpretations can legitimately be attached to the results. Measurement must therefore occupy the gatekeeper position. Adding measurement while leaving the inherited analytical architecture untouched would reproduce the same inversion under a superficially expanded curriculum.

The importance of this failure extends far beyond any individual university, professional organization or national HTA system. Measurement inversion has become a global phenomenon. The same analytical architecture has been transmitted internationally: health-state valuation, preference elicitation, utilities, QALYs, incremental QALYs, cost-per-QALY ICERs, reference cases, decision models, simulation and uncertainty analysis. Countries differ in the precise methods they emphasize and in the institutional roles assigned to HTA, but the underlying intellectual sequence has been remarkably persistent. Downstream analytical methods are taught, standardized and applied while the logically prior requirements of representational measurement, ratio measurement, dimensional homogeneity and admissible arithmetic receive little or no corresponding attention.

The persistence of this pattern can be understood as the international transmission of a methodological memeplex: a mutually reinforcing collection of concepts, methods, assumptions and institutional practices that is reproduced as an integrated intellectual package. Utilities support QALYs; QALYs support incremental QALYs; incremental QALYs support cost-per-QALY ICERs; ICERs support reference-case decision models; guidelines standardize those models; journals publish them; professional organizations teach them; universities incorporate them into curricula; agencies and payers request them; and employers expect graduates to know how to produce them. Each component reinforces the apparent legitimacy of the others. The framework becomes credible because it is everywhere, and it becomes everywhere because it is treated as credible.

This transmission mechanism also explains why measurement failure can survive for generations. Students are taught an accepted analytical framework and become analysts, researchers and faculty who reproduce it. Professional guidelines formalize the same methods. Journals require conformity with established analytical conventions. Research centers train another generation in their application. International professional networks disseminate them across jurisdictions. Institutional replication is then easily mistaken for scientific replication. Yet what has been repeatedly reproduced is the memeplex, not the measurement proof required to establish the legitimacy of its central quantitative constructs.

Global acceptance therefore provides no defense against the measurement challenge. A QALY does not acquire ratio properties because it is taught on several continents. An ICER does not acquire scientific meaning because multiple agencies request it. A reference case does not rescue an inadmissible quantity by requiring everyone to use it in the same way. Replication of a methodology is not replication of its measurement foundation. Once that distinction is recognized, the extraordinary international success of legacy HTA becomes evidence of the scale of the transmission problem rather than evidence of the scientific legitimacy of the framework transmitted.

Once students understand that measurement precedes arithmetic, that attributes must be identified, that manifest and latent attributes require different measurement pathways, that scale properties constrain arithmetic, that dimensional homogeneity constrains quantitative relationships and that proportional claims require appropriate ratio structures, they will inevitably apply those standards to the methods they are subsequently taught. The contradiction will then be obvious.

A curriculum cannot coherently teach measurement first and legacy HTA second without requiring legacy HTA to pass the measurement standards it previously ignored. Once that test is imposed, the central QALY-based architecture fails. QALYs cannot simply continue to be taught as measures; incremental QALYs cannot be assumed to be legitimate quantities; and cost-per-QALY ICERs cannot retain scientific meaning merely because they remain components of established economic-evaluation practice.

The problem therefore represents more than a curriculum imbalance. It identifies the knowledge failure that permitted measurement inversion to become accepted HTA practice and subsequently to be transmitted internationally as an apparently validated methodological framework. The appropriate response is not curriculum supplementation. It is curriculum reconstruction.

That reconstruction begins with the definition of the attribute. It distinguishes manifest from latent attributes and recognizes their different measurement requirements. Manifest attributes require appropriate linear ratio measurement. Latent attributes require an appropriate measurement model capable of constructing a unidimensional measurement structure. Scale properties must be established before arithmetic is permitted. Quantitative therapy-impact claims must then be empirically evaluable, replicable and capable of falsification.

This transition to lawful measurement is not an optional enrichment of existing HTA education. It is the only scientifically defensible response to the failure of legacy HTA. It provides students with the foundations required to understand the standards of ratio measurement for manifest attributes and the fundamentally different requirements for measuring latent attributes. Measurement once again becomes the gatekeeper for arithmetic and quantitative claims. After more than forty years of measurement inversion, HTA can return to the standards of normal science.

## THE TRANSFORMATION TO LAWFUL MEASUREMENT

Abandonment of the legacy HTA curriculum and the QALY does not leave universities, research centers or HTA programs without a way forward. As noted above, the required transformation has already been set out in a nine-unit program, *Reconstructing HTA Programs: Measurement before Arithmetic* [7] . Its importance is not that it offers another variant of conventional HTA education. It does the opposite. It abandons the inherited sequence and reconstructs HTA from the measurement foundations that should have preceded quantitative analysis from the outset 40 years ago. Utilities, QALYs and reference-case simulations have no part to play.

The defense of legacy HTA that an analytical framework was required to fill evidence gaps at the time of marketing approval fundamentally mistakes the purpose of science. The objective is not to manufacture numerical answers where evidence is absent. It is to generate evidence through lawful measurement and empirical testing. Representational measurement provides the foundation for that process: define the attribute, construct an appropriate measure, formulate a therapy-impact claim and subject that claim to empirical evaluation, replication and potential falsification. Legacy HTA reverses this sequence. It responds to missing evidence by constructing assumptions, utilities, QALYs and simulated outcomes and then treating the resulting numbers as though they constituted evidence. They do not. Numerical output from a model cannot substitute for measurement of the attribute about which the claim is being made.

Evidence gaps are real, but their existence does not authorize imaginary claims. Where evidence required for a decision is missing, the scientifically defensible response is to identify the missing therapy-impact claims and establish protocols capable of generating the required evidence using lawful measures. Uncertainty should be acknowledged as uncertainty until evidence resolves it. The necessity of making a decision cannot transform an unmeasured quantity into a measured one or a simulated claim into an empirical fact.

This distinction reaches directly back to the scientific revolution. The Royal Society expressed the principle in its motto, Nullius in verba: “take nobody’s word for it.” [8] Scientific claims must answer to observation, measurement and demonstration rather than authority or convention. Legacy HTA asks for precisely the reverse: accept constructed numerical claims because an accepted framework, reference case or professional consensus has generated them. Normal science demands evidence. If the evidence does not exist, the task is to create it through measurement and empirical inquiry not to create numbers in its place.

The transformation begins with a simple reversal: measurement must precede arithmetic. Students are not introduced first to utilities, QALYs, ICERs and reference-case models and subsequently asked to consider their measurement properties. They begin by identifying the attribute about which a therapy-impact claim is to be made and determining what would constitute a lawful measure of that attribute. Only after the measurement requirements have been satisfied does arithmetic become admissible.

The nine-unit program develops this reconstruction progressively. It begins with measurement inversion and the consequences of allowing arithmetic to precede measurement. It then establishes the importance of attributes, scale types, admissible transformations, dimensional homogeneity and the restrictions that measurement properties impose upon quantitative claims.

The program then moves from measurement principles to their application in therapy evaluation. Students learn why assigning numbers to observations does not itself create measurement; why conventional multiattribute preference scores cannot perform the proportional measurement role demanded of them by the QALY; and why arithmetic cannot repair a measurement failure that has already occurred. Therapy-impact claims must instead be constructed from measures whose properties have been established and must be capable of empirical evaluation, replication and falsification.

This changes the intellectual structure of HTA. The reference case is no longer the organizing principle. Measurement is. The question is no longer whether an analysis conforms to an accepted methodological convention but whether the quantities entering the analysis are lawful measures and whether the operations performed upon them are permitted by their measurement properties.

The nine units are therefore not intended to repair legacy HTA. They provide the basis for its replacement. Utilities, QALYs and cost-per-QALY reference-case claims are not retained and surrounded by additional qualifications. Constructs that fail the measurement gate are abandoned. Those that survive lawful measurement scrutiny can remain. The curriculum is reconstructed from what can legitimately be measured rather than from what legacy HTA has traditionally chosen to calculate.

The new curriculum presents a much more straightforward framework for therapy-impact assessment. The starting point is the attribute about which the therapy-impact claim is to be made. If the selected attribute is manifest, such as survival time, hospital days, walking distance or treatment administrations, it must be represented by an appropriate linear ratio measure, with equal units, a non-arbitrary zero and the admissible transformations required for meaningful proportional comparisons.

If the selected attribute is latent, such as pain, fatigue, physical function or need fulfilment, a different measurement pathway is required. The Rasch measurement model provides a framework for constructing a unidimensional latent continuum through person-item conjoint measurement, invariance and specific objectivity. The resulting logit scale provides an additive representation of the latent attribute, while exponentiation permits multiplicative comparisons to be expressed through the corresponding odds. The manifest and latent pathways are therefore different, but both require measurement to be established before quantitative therapy-impact claims are made.

The framework is remarkably simple: identify the attribute, determine whether it is manifest or latent, establish the appropriate lawful measure, determine the arithmetic permitted by that measure, and only then proceed to the assessment of therapy impact. There is no requirement for utilities, QALYs or composite preference indices to intervene between the observed treatment effect and the attribute about which the claim is being made.

This transformation does not require universities, research centers or HTA programs to wait for an HTA agency, professional organization or other central authority to authorize compliance with measurement standards. The standards of lawful measurement already exist. Nor is there a requirement to wait for another generation of methodological debate. Once a quantitative operation has been shown to lack the measurement properties necessary for its claimed interpretation, institutional convention cannot preserve its scientific legitimacy.

This point is particularly important given the global transmission of legacy HTA described in the preceding section. The methodological memeplex has been reproduced internationally through universities, professional organizations, journals, guidelines, research centers, agencies, payers and industry. That history may make institutional transition difficult, but it does not create a requirement for coordinated global abandonment. The same decentralized mechanisms that transmitted measurement inversion can now facilitate its replacement. Individual universities, research groups and programs can reconstruct their curricula without waiting for the entire HTA community to move simultaneously.

Indeed, no central authority is required. Representational measurement is not a competing HTA doctrine requiring institutional endorsement. The requirements that an attribute be defined, that measurement properties be demonstrated, that arithmetic be admissible for the scale employed and that quantitative claims possess empirical meaning are logically prior to the methodological conventions of HTA. An institution does not require permission to insist that measurement precede arithmetic.

The significance of the accumulated measurement critique should therefore not be understated. HTA is not being asked to abandon an established curriculum merely because a competing methodological school has appeared. The issue is logically prior to methodological preference. Once measurement becomes the gatekeeper, every quantitative construct must demonstrate that it possesses the properties required for the claims made with it. If a construct passes that gate, there is nothing to abandon. If it fails the gate, there is nothing to rescue. The QALY fails.

The consequences extend beyond removal of a single construct. The curriculum that produced and sustained the QALY was itself organized around measurement inversion. Health-state valuation, utility algorithms, QALYs, incremental QALYs, cost-per-QALY ICERs and reference-case simulation formed a mutually reinforcing analytical structure in which increasingly complex arithmetic obscured the absence of the logically prior measurement foundation. Abandoning the QALY while preserving that intellectual sequence would therefore leave the underlying failure intact.

Curriculum reconstruction must instead break the transmission mechanism described in the preceding section. Faculty must be provided with the measurement framework that may have been absent from their own education, while students must acquire the intellectual tools required to distinguish numerical calculation from scientifically defensible measurement. The objective is not merely to replace one set of analytical techniques with another. It is to prevent another generation from inheriting the same methodological memeplex without being equipped to interrogate its measurement foundations.

The nine-unit program demonstrates that this transition can begin immediately. It provides a coherent structure for replacing measurement inversion with measurement discipline. It also demonstrates that abandoning legacy HTA does not create an analytical vacuum. What disappears are constructs that cannot survive the measurement gate. What replaces them is a framework in which therapy-impact assessment begins with attributes that can be defined and measured and proceeds only through quantitative operations justified by the properties of those measures.

This is a fundamental change in the purpose of HTA education. Technical competence in established analytical procedures is no longer sufficient. Students should be able to ask what attribute is being claimed, how it has been measured, what scale has been constructed, what transformations preserve its meaning, what arithmetic is admissible, whether dimensional requirements have been satisfied and whether the resulting therapy-impact claim can be empirically evaluated, replicated and potentially falsified.

The transition is not a choice between competing methodologies or alternative schools of economic evaluation. Once the requirements of representational measurement are imposed, there is only one scientifically credible framework for assessing therapy impact: define the attribute, establish a lawful measure of that attribute, determine the arithmetic permitted by the properties of that measure, and construct an empirically evaluable therapy-impact claim. Legacy HTA cannot stand alongside this framework as an alternative scientific methodology. Its central constructs were created by allowing arithmetic to precede measurement, and those that fail the measurement gate have no claim to quantitative legitimacy. The task is therefore not to choose between legacy HTA and lawful HTA, nor to combine elements of each. It is to abandon the failed numerical conventions of legacy HTA and reconstruct therapy-impact assessment from lawful measurement.


## 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.


## ABOUT THE AUTHOR

Dr Paul Langley is a health economist. Originally from the UK, Dr Langley completed his graduate studies in Canada and has taught in the UK, Canada, Australia and the United States. After over 20 years teaching university in Australia, Dr Langley moved to take up academic positions in the United States. He now lives in Tucson, Arizona to enjoy the Catalina mountains, the Sonoran desert, the absence of mosquitos and his swimming pool.

## REFERENCES


[1] Stevens S. On the Theory of Scales of Measurement. *Science*. 1946;103(2684):677–680.

[2] Krantz DH, Luce RD, Suppes P, Tversky A. *Foundations of Measurement. Vol. 1: Additive and Polynomial Representations*. New York: Academic Press; 1971

[3] Maimon Research: Global Catalog of Measurement Inversion. https://maimonresearch.com/ai-llm-usa-uk/#united-states

[4] Langley P. Nonsense on Stilts: Why the QALY never existed as a lawful measure. Logit Working Paper No. 845. https://maimonresearch.com/logit-working-paper-no-845-september-2026/

[5] Langley P. United States: Representational measurement – The missing scientific foundation of the national HTA/HEOR knowledge base. Maimon Working Paper No 1908. https://maimonresearch.com/logit-working-paper-no-1908-july-2026/

[6] Drummond M, Sculpher M, Claxton K et al. Methods for the Economic Evaluation of Health Care Programmes (4th Ed). New York: Oxford University Press, 2015

[7] Maimon Research. *Reconstructing HTA Programs: Measurement before Arithmetic*. https://maimonresearch.com/hta-reconstruction-program-and-fees/

[8] Wootton D. The Invention of Science: A new history of the scientific revolution, New York: HarperCollins, 2015