

MAIMON RESEARCH LLC

**ARTIFICIAL INTELLIGENCE LARGE LANGUAGE
MODEL INTERROGATION**



**REPRESENTATIONAL MEASUREMENT FAILURE IN
HEALTH TECHNOLOGY ASSESSMENT**

THE IMPOSSIBLE QALY: THE PROOF

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

LOGIT WORKING PAPER No 8080 OCTOBER 2026

www.maimonresearch.com

Tucson AZ

ABSTRACT

Background: For more than four decades, the quality-adjusted life year (QALY) has provided the principal outcome measure for cost-utility analysis in health technology assessment (HTA). Its conventional construction, $Q = U \times T$, assumes that a health-state preference or utility value, U , can function as a proportional adjustment factor for duration, T . This paper challenges that assumption by applying the requirements of representational measurement, including scale properties, commensurability, dimensional homogeneity and admissible arithmetic.

Methods: The QALY construction is evaluated through a measurement-first sequence: Attribute $\rightarrow$ Measurement $\rightarrow$ Scale properties $\rightarrow$ Admissible transformations $\rightarrow$ Admissible arithmetic $\rightarrow$ Quantitative claim. The analysis distinguishes dimensional cancellation from commensurability and establishes the requirements for a lawful proportional adjustment. A time proportion, $T1 / T2$, has empirical meaning when $T1$ and $T2$ are commensurable ratio measures of duration. The same requirements are applied to preference values and to the proposed mapping from preference space to time space.

Results: The QALY fails two independent measurement requirements. First, U is not established as a lawful ratio between commensurable measures of a defined preference attribute. Consequently, $U = 0.6$ does not establish $P1 / P2 = 0.6$. Second, even if a lawful preference ratio could be constructed, no measurement relationship establishes that $P1 / P2 = 0.6$ maps to $T1 / T2 = 0.6$. Numerical equality and dimensional cancellation cannot establish equivalence between proportions originating in different measurement spaces. The QALY therefore cannot exist as a lawful measure of quality-adjusted time. The failure originates upstream in health-state valuation and propagates through utilities, incremental QALYs, cost-per-QALY claims, decision models and QALY-dependent reference cases.

Conclusions: The QALY is the most visible manifestation of a broader failure in legacy HTA: arithmetic has been allowed to precede measurement. Its removal therefore requires reconstruction rather than methodological repair. HTA education and research should begin with defined attributes and recognize two measurement pathways for therapy-impact claims: linear ratio measurement for manifest attributes and Rasch measurement for latent attributes. Scale properties must determine admissible arithmetic, with quantitative claims subject to empirical evaluation, replication and falsification. Legacy HTA cannot be preserved as an alternative quantitative paradigm once its foundational measurement failure is recognized. The required transition is from procedural competence in inherited numerical techniques to measurement competence: attribute before number, measurement before arithmetic, and lawful measurement before quantitative claim.

INTRODUCTION

For more than four decades, the quality-adjusted life year (QALY) has occupied a central position in health technology assessment (HTA). It is used to represent therapy impact, compare interventions, construct incremental outcomes and provide the denominator for cost-utility claims. Yet one logically prior question has received remarkably little attention: is the QALY a measure? The fact that a numerical value can be calculated does not establish that it represents a measured quantity. Before arithmetic is undertaken, the measurement properties required to support that arithmetic must first be demonstrated.

Representational measurement imposes a required sequence on quantitative analysis: Attribute → Measurement → Scale properties → Admissible transformations → Admissible arithmetic → Quantitative claim. Numbers acquire quantitative meaning from the empirical relationships they represent, not from the mathematical operations subsequently performed upon them^{1 2 3}. Different measurement scales therefore permit different operations. In particular, proportional interpretation requires ratio measurement, with an invariant unit and a non-arbitrary zero. Multiplication cannot be justified merely because numerical values have been assigned to the quantities involved.

The conventional QALY reverses this sequence. Its construction is expressed as $Q = U \times T$, where Q is quality-adjusted time, U is a health-state preference or utility value and T is duration. Duration is a ratio measure in time space. The role assigned to U is more demanding than is usually acknowledged. If $U = 0.6$, the QALY construction treats that value as a proportional adjustment to duration: one year is reduced to 0.6 quality-adjusted years. The critical measurement question is therefore not whether 0.6 can be multiplied by time. Arithmetic makes that trivial. The question is whether $U = 0.6$ has been established as a lawful proportion and, if so, what warrants interpreting that proportion as a 60% adjustment to duration.

The purpose of this paper is to provide a direct measurement proof that the conventional QALY cannot satisfy these requirements and therefore cannot exist as a lawful measure of quality-adjusted time. The proof does not depend upon a particular national value set, preference instrument, reimbursement agency or HTA reference case. It follows from the measurement requirements imposed by the quantitative claim itself and consequently applies wherever the QALY construction is used. The heritage of the QALY is decades of quantitative claims failure.

The argument proceeds in five stages. First, commensurability establishes the conditions under which a ratio possesses empirical proportional meaning. Second, dimensional homogeneity distinguishes dimensional cancellation from measurement meaning. Third, proportional adjustment establishes what is required for a numerical factor to lawfully adjust a ratio measure such as time. Fourth, these requirements are applied directly to $Q = U \times T$ to demonstrate the two independent failures that make the QALY impossible. Finally, the implications are considered for incremental QALYs, cost-per-QALY claims, economic modelling and QALY-based HTA. The issue is not whether the QALY is useful or widely accepted. The logically prior question is why it cannot exist as the measure it claims to be.

1. COMMENSURABILITY

Commensurability is fundamental to meaningful proportional comparison. Two quantities are commensurable when they are measures of the same attribute within the same measurement space and can therefore be compared through a common measurement structure. For a ratio to express how much of one quantity is represented by another, the numerator and denominator must be commensurable ratio measures. Arithmetic division alone cannot establish this relationship. Any two numbers can be divided, but the numerical result does not necessarily represent an empirically meaningful proportion.

Duration provides a straightforward example. Consider two measured durations, T1 and T2, where T1 is six months and T2 is ten months. Their ratio is:

$$T1 / T2 = 6 \text{ months} / 10 \text{ months} = 0.6$$

The interpretation is unambiguous. Six months is 60% of ten months. This statement is meaningful because T1 and T2 are commensurable measures of the same attribute, duration. They share an invariant measurement unit and a common non-arbitrary zero. The ratio is also invariant to a permissible change in units. Expressing the durations as 0.5 years and 0.8333 years gives the same ratio:

$$0.5 / 0.8333 = 0.6$$

The number 0.6 therefore has empirical proportional meaning. It is not simply the numerical result of an arithmetic operation.

This distinction becomes critical when numbers generated in different measurement spaces are compared. Suppose a ratio of two measures of another attribute also has the numerical value 0.6. Numerical equality does not establish empirical equivalence. A ratio describing two quantities of duration and a ratio describing two quantities of another attribute may both equal 0.6 while expressing entirely different empirical relationships. The meaning of each ratio derives from the commensurable quantities from which it was constructed, not from the fact that the arithmetic produces the same number.

This is why a numerical value cannot simply be declared to represent a proportion. If a value of 0.6 is claimed to mean 60% of some attribute, it must be possible to identify the two commensurable ratio measures from which that proportion is derived. If P represents a preference attribute, for example, the proportional claim requires:

$$P1 / P2 = 0.6$$

where P1 and P2 are commensurable ratio measures of that same preference attribute. Only then can 0.6 legitimately mean that P1 represents 60% of P2.

Commensurability therefore supplies the empirical foundation for proportional interpretation. The number 0.6 does not carry the meaning “60%” simply because it lies between zero and one. It

acquires that meaning only through a demonstrated ratio between commensurable measures of the same attribute. This requirement becomes decisive when a proportion generated in one measurement space is subsequently claimed to provide a proportional adjustment to a quantity measured in another.

2. DIMENSIONAL HOMOGENEITY

Dimensional homogeneity is a necessary requirement of meaningful quantitative analysis. Every term in a quantitative expression must be consistent with the dimensional character of the quantities involved. Quantities that are added or subtracted must belong to the same dimensional class, while multiplication and division generate dimensions determined by those of the quantities being combined. An equation that violates dimensional homogeneity cannot represent a coherent quantitative relationship. Dimensional consistency must therefore be established before the resulting arithmetic can be given an empirical interpretation.

Duration provides a simple example. Let T_1 and T_2 represent two ratio measures of duration. Both belong to the dimensional class $[T]$, where $[T]$ denotes time. If T_1 is divided by T_2 :

$$A_t = T_1 / T_2$$

the dimensions are:

$$[A_t] = [T] / [T] = 1$$

The dimensions cancel and A_t is dimensionless. If $T_1 = 6$ months and $T_2 = 10$ months:

$$A_t = 6 / 10 = 0.6$$

The resulting factor can lawfully express a proportional relationship because the dimensions cancel and, critically, T_1 and T_2 are commensurable ratio measures of the same attribute. Dimensional homogeneity and commensurability are therefore complementary requirements, but they perform different functions.

This distinction is essential. Dimensional cancellation by itself does not establish proportional meaning. Suppose P_1 and P_2 were ratio measures of a preference attribute belonging to dimensional class $[P]$. Their ratio would similarly produce:

$$A_p = P_1 / P_2$$

and:

$$[A_p] = [P] / [P] = 1$$

A_p would also be dimensionless. If A_p happened to equal 0.6, this would establish a 60% relationship within preference space, provided P_1 and P_2 were commensurable ratio measures. It would not establish that this preference ratio was equivalent to the time ratio $A_t = 0.6$.

Consequently:

$$[P] / [P] = 1$$

and:

$$[T] / [T] = 1$$

do not establish:

$$P1 / P2 = T1 / T2$$

Two ratios can be dimensionless and numerically identical while representing entirely different empirical relationships. Their dimensions have cancelled, but their measurement origins have not disappeared. The meaning of each ratio remains determined by the quantities and attributes from which it was constructed.

The distinction can therefore be stated simply: dimensional cancellation is an algebraic result; commensurability is a measurement requirement. Dimensional homogeneity may establish that an equation is dimensionally permissible, but it cannot demonstrate that a dimensionless factor derived in one measurement space has the empirical meaning required to adjust a quantity in another. If such an adjustment is proposed, a separate measurement relationship between the two spaces must be established. This requirement becomes critical when a health-state preference value is used as a purported proportional adjustment to duration.

3. PROPORTIONAL ADJUSTMENT

A proportional adjustment changes the magnitude of a measured quantity while preserving a defined relationship to that quantity. Where the quantity being adjusted is measured on a ratio scale, the adjustment factor must itself possess a lawful proportional interpretation. It is not sufficient for the proposed adjustment factor merely to be numerical, to lie between zero and one, or to be dimensionless. Its interpretation as a proportion must first be established through measurement.

Duration again provides the clearest example. Suppose T2 represents an original duration of ten months and T1 represents an adjusted duration of six months. The proportional adjustment factor is:

$$At = T1 / T2$$

Therefore:

$$At = 6 \text{ months} / 10 \text{ months} = 0.6$$

The adjusted duration can then be recovered as:

$$T1 = A_t \times T2$$

or:

$$6 \text{ months} = 0.6 \times 10 \text{ months}$$

The multiplication is meaningful because $A_t = 0.6$ was derived from two commensurable ratio measures of the same attribute, duration. The adjustment factor therefore has an established empirical interpretation: $T1$ is 60% of $T2$. The number 0.6 does not acquire this meaning from multiplication. It possessed proportional meaning before the multiplication was performed.

Now suppose an adjustment factor is generated from another measurement space. Let $P1$ and $P2$ represent two lawful ratio measures of the same preference attribute. Their ratio would be:

$$A_p = P1 / P2$$

If:

$$A_p = 0.6$$

then $P1$ is 60% of $P2$ within preference space. This would establish a lawful proportional relationship between $P1$ and $P2$. It would not, however, establish that duration should also be reduced to 60% of its original value. The preference ratio and the time ratio have different empirical origins.

To use A_p as an adjustment factor for duration, an additional relationship must therefore be demonstrated:

$$P1 / P2 \rightarrow T1 / T2$$

If:

$$P1 / P2 = 0.6$$

it must be established independently that this relationship implies:

$$T1 / T2 = 0.6$$

The numerical equality of the two ratios cannot provide that proof. Nor can the fact that both ratios are dimensionless. A 60% relationship between two measured amounts of one attribute does not automatically become a 60% relationship between two measured amounts of another attribute.

This establishes the measurement requirements for proportional adjustment. A factor applied directly within a ratio measurement space can be justified when it is derived from commensurable ratio measures of the attribute being adjusted. A factor originating in another measurement space requires two demonstrations: first, that it is itself a lawful proportion constructed from

commensurable ratio measures of a defined attribute; and second, that a measurement relationship maps that proportion to the required proportion in the measurement space being adjusted. Without both demonstrations, multiplication can still be performed numerically, but the resulting product has no established interpretation as a proportionally adjusted measure.

4. THE PROOF FOR AN IMPOSSIBLE QALY

The conventional QALY claims to create a measure of quality-adjusted time by multiplying a health-state preference or utility value, U , by duration, T :

$$Q = U \times T$$

The interpretation assigned to U is explicitly proportional. If $U = 1$, duration is left unchanged. If $U = 0.6$, one year in the health state is treated as 0.6 quality-adjusted years. If $U = 0.5$, one year is treated as one-half of a quality-adjusted year. U is therefore not functioning merely as a numerical description of a health state. It is being used as a proportional adjustment factor for time. The requirements established in the preceding sections can now be applied directly to this claim.

The first requirement is that U must itself possess lawful proportional meaning. If $U = 0.6$ is to mean 60% within preference space, there must be a defined preference attribute, P , measured on a ratio scale, together with two commensurable measures $P1$ and $P2$ such that:

$$Ap = P1 / P2 = 0.6$$

Only then would it be legitimate to state that $P1$ represents 60% of $P2$. Conventional health-state preference systems do not establish this measurement relationship. A health state assigned $U = 0.6$ is located numerically within a preference-value system, commonly anchored by values such as 1 for full health and 0 for death. Those numerical reference points do not establish a non-arbitrary zero, an invariant unit, unidimensionality, commensurable ratio measures or the legitimacy of proportional comparisons. Consequently:

$$U = 0.6$$

does not establish:

$$P1 / P2 = 0.6$$

This is the first failure. U has not been demonstrated to be the lawful proportion required for proportional adjustment.

The QALY then encounters a second and independent failure. Suppose, purely for argument, that a preference attribute could be measured on a lawful ratio scale and that:

$$P1 / P2 = 0.6$$

had been established. This would demonstrate that P1 is 60% of P2 within preference space. It would still provide no measurement warrant for reducing duration to 60% of its original value. The required proportional relationship within time space remains:

$$At = T1 / T2 = 0.6$$

A separate measurement relationship must therefore demonstrate:

$$P1 / P2 \rightarrow T1 / T2$$

The conventional QALY provides no such mapping. Instead, it takes the numerical value $U = 0.6$ from preference space and applies it directly to duration:

$$Q = 0.6 \times T$$

The construction therefore assumes precisely the relationship that measurement must first demonstrate. The fact that a preference ratio and a time ratio might both be dimensionless does not establish their equivalence. Nor does their numerical equality. A ratio of 0.6 within one measurement space cannot simply be declared to represent a ratio of 0.6 within another.

The proof therefore consists of two independent failures:

Failure 1: U is not demonstrated to be a lawful preference proportion.

$$U = 0.6 \text{ does not establish } P1 / P2 = 0.6$$

Failure 2: Even if a lawful preference proportion existed, there is no demonstrated mapping from preference proportion to time proportion.

$$P1 / P2 = 0.6 \text{ does not establish } T1 / T2 = 0.6$$

Failure at either stage is sufficient to invalidate the QALY. The conventional construction fails at both.

The conclusion follows directly. $Q = U \times T$ can be performed arithmetically, but its product cannot be interpreted as a lawful measure of quality-adjusted time. U has not been established as the required proportional measure, and even granting such a measure cannot establish the required mapping from preference space to time space. The QALY is therefore an impossible measurement construct. There are no missing measurement properties waiting to be demonstrated. The failure is inherent in its construction.

The implications extend to the terminology of HTA itself. The term QALY should be excised from the lexicon of health technology assessment. It is not simply that the QALY has been demonstrated to be a failed measurement construct, as though it were one unsuccessful measure that might be repaired, recalibrated or replaced by an improved version of itself. The more fundamental conclusion is that there is no lawful quantitative object corresponding to what the term claims to

represent. There is duration, which can be measured in time space, and there are numerical representations of health-state preferences, but their multiplication does not create a new measure called quality-adjusted time. Once the arithmetic convention is shown to violate the requirements of measurement, there is nothing left to preserve under the name. The QALY is a will-o'-the-wisp: an apparent quantitative object that has led generations of HTA practitioners, researchers and students along a measurement path that never existed. It belongs not in the continuing vocabulary of quantitative HTA but as a historical footnote to the consequences of allowing arithmetic to precede measurement.

5. THE TIP OF THE ICEBERG

The failure of the QALY is only the tip of the iceberg. It is not an isolated defect appearing at the final stage of an otherwise lawful measurement process. It is the inevitable consequence of a more fundamental failure that begins upstream with the valuation of health states. Legacy HTA reversed the required scientific sequence. Instead of proceeding from attribute to measurement, from measurement to scale properties, and from scale properties to admissible arithmetic, it began by assigning numerical values to health states and subsequently treated those values as though their measurement properties had already been established. Arithmetic preceded measurement.

The time trade-off illustrates the problem. A respondent is typically asked to consider time in an impaired health state and a shorter period in full health, with an indifference point used to generate a numerical health-state value. Time itself is a ratio measure, but the fact that time enters the elicitation procedure does not make the resulting health-state value a ratio measure of health, quality of life or preference. The procedure must still identify a single attribute, demonstrate unidimensionality, establish the empirical structure represented by the numerical values, provide a non-arbitrary zero and invariant unit, and demonstrate the admissible transformations and arithmetic supported by the resulting scale. Taking a ratio of durations within an elicitation procedure cannot transfer the ratio properties of time to a different attribute.

The problem compounds as the process continues. Multidimensional descriptive systems such as EQ-5D define health states across several attributes. Preferences are elicited for those states, statistical models estimate weights or decrements, and algorithms generate numerical index values. Successive numerical transformations may be statistically sophisticated, reproducible and useful for describing preferences, but none of these operations can manufacture measurement properties that were absent at the preceding stage. Assigning numbers, ranking states, estimating coefficients, summing weighted components or anchoring values at zero and one does not establish a unidimensional ratio scale. Yet the resulting values are subsequently called utilities and treated as though ratio arithmetic were admissible.

The QALY therefore inherits an upstream measurement failure before the multiplication $Q = U \times T$ is even attempted. The proposed utility value has already passed through a sequence in which valuation has been treated as measurement and numerical manipulation has substituted for demonstration of scale properties. The QALY then adds the further failures established in the preceding section: U is not demonstrated to be a lawful proportional measure, and no mapping demonstrates that a proportion in preference space becomes the corresponding proportion in time

space. The QALY does not create the measurement failure. It exposes the cumulative consequences of a measurement failure that began with health-state valuation.

The consequences then propagate downstream. Incremental QALYs cannot acquire measurement properties absent from the QALYs from which they are calculated. A cost-per-QALY ratio cannot acquire a lawful outcome denominator merely because monetary cost is divided by a numerical QALY difference. Decision models cannot manufacture measurement properties through simulation, extrapolation or probabilistic analysis. Most importantly, an HTA reference case built around QALY-based cost-utility analysis cannot escape the failure by prescribing the methods through which QALYs should be generated and applied. A reference case can standardize a procedure; it cannot transform an inadmissible numerical construction into a measure. Once the QALY disappears as a lawful measure, the quantitative foundation of any reference case that requires QALY-based cost-utility claims disappears with it. The iceberg is therefore much larger than the QALY. The underlying problem is the legacy HTA commitment to arithmetic before measurement.

The consequences of this proof are unsettling because they are retrospective as well as prospective. Over four decades, many thousands of therapy-impact claims have been constructed from utilities, QALYs, incremental QALYs and cost-per-QALY ratios and submitted to HTA agencies as quantitative evidence. Professional organizations developed methodological standards around these constructs, while agencies incorporated them into reference cases and decision processes. If the logically prior requirements of lawful measurement were never satisfied, the problem cannot be confined to future methodological guidance. It reaches backward into the accumulated evidentiary record. Every QALY-dependent claim must confront the same question: what was actually measured? Neither widespread acceptance, methodological standardization nor institutional authority can retrospectively supply measurement properties that the underlying construct does not possess. This is the QALY heritage: decades of measurement failure.

The law of unintended consequences is therefore formidable. A framework intended to make therapy assessment more rigorous has generated an enormous body of apparently quantitative evidence whose central outcome construct cannot exist as a lawful measure. Organizations such as ISPOR, NICE and PBAC consequently face a question extending far beyond whether future methods should be revised: what status can be accorded to past recommendations, reimbursement assessments and cost-effectiveness claims whose quantitative justification depended upon a QALY that cannot exist? The issue is not whether those QALY claims were right or wrong. That would presume the existence of a legitimate quantitative claim capable of being evaluated. There was none. The arithmetic produced a number, but that number represented nothing. It had no demonstrated measurement referent. The unintended consequence is an accumulated global HTA evidence base containing thousands of QALY-dependent claims whose supposed quantitative foundation disappears once the requirements of lawful measurement are imposed.

6. THE DEMISE OF LEGACY HTA

For more than forty years, thousands of researchers, academics, consultants, agency analysts and students have accepted the foundational assumptions of legacy HTA substantially at face value. Health states could be described across multiple attributes, preferences could be elicited for those

states, numerical values could be assigned, those values could be designated utilities, and utilities could be multiplied by time to create QALYs. Once these assumptions became established, research concentrated overwhelmingly on improving the procedures operating within the framework rather than asking whether the framework itself satisfied the requirements of measurement.

The extraordinary feature of this history is the absence of the logically prior questions. What attribute is being measured? Is it unidimensional? What empirical relational structure is represented by the numerical values? What scale of measurement has been constructed? Does it possess an invariant unit and a non-arbitrary zero? What transformations are admissible? What arithmetic does the scale permit? Can a multidimensional health-state construct satisfy the axioms required for measurement of a single attribute? These are not optional philosophical questions. They determine whether quantitative operations have empirical meaning. Yet they appear to have remained largely outside the intellectual boundaries within which legacy HTA developed.

This intellectual isolation became self-reinforcing. Utilities supported QALYs; QALYs supported cost-utility analysis; cost-utility analysis generated cost-per-QALY claims; reference cases institutionalized these methods; agencies required them; universities taught them because agencies required them; journals published studies employing them; professional organizations codified them as accepted practice; and their widespread use was then treated as evidence of legitimacy. This can be described as an epistemic memplex: a mutually reinforcing system of assumptions, methods, institutions and educational practices whose individual components derive credibility from the other components of the same system. No conspiracy or deliberate deception is required. Indeed, the system is more durable precisely because those operating within it may have no reason to recognize that its foundational assumptions require external measurement proof.

The result is extraordinary. A subject with explicitly quantitative pretensions and global reach has developed in which apparent numerical sophistication has been allowed to substitute for lawful measurement. Preference elicitation, statistical estimation, mapping, decision modelling, simulation, sensitivity analysis and probabilistic analysis can all be performed with considerable technical sophistication while leaving unanswered the logically prior question: what has been measured? Arithmetic has effectively been permitted to authorize itself. Yet no amount of mathematical sophistication can transform a numerical representation into a measure if the required measurement properties are absent.

The consequences of exposing this failure will be profound. The QALY proof does not identify a minor methodological defect requiring another adjustment to existing guidance. It exposes a foundational error extending from health-state valuation through utilities and QALYs to cost-utility claims and QALY-dependent reference cases. Once the requirements of representational measurement and the constraints imposed by scales of measurement are brought inside the intellectual boundaries of HTA, much of what has been accepted as quantitative measurement can no longer retain that status. An entire subject area faces the fact that its central quantitative framework has been constructed upon false measurement assumptions.

This is the demise of legacy HTA. It does not mean the demise of health technology assessment or quantitative evaluation. It means the end of a particular intellectual framework in which

arithmetic was permitted to precede measurement. The alternative is reconstruction: attributes before numbers, measurement before arithmetic, scale properties before mathematical operations, and empirically evaluable therapy-impact claims before modelling. The failure of legacy HTA is therefore not that it lacked numbers. It is that, for forty years, it mistook numbers and arithmetic for measurement.

7. RECONSTRUCTING HEALTH TECHNOLOGY ASSESSMENT

The failure of legacy HTA has an immediate implication for education. There is no valid reason for universities, research centers, professional organizations and other groups to continue teaching an HTA curriculum organized around methods whose quantitative foundations fail the requirements of lawful measurement. Indeed, the opposite conclusion follows. Now arithmetic-before-measurement is recognized as the foundational error, continuing to reproduce the same sequence for another generation of students is impossible to justify.

The transition is, paradoxically, made easier by the present structure of HTA education. Representational measurement does not have to displace an established competing measurement framework. Across the program-content assessments undertaken to date, there is little evidence that the requirements of lawful measurement occupy any foundational position in HTA education. Students are taught economic evaluation, utilities, QALYs, ICERs, decision models, uncertainty analysis and related techniques, but the logically prior questions are largely absent: What is the attribute? How is it measured? What scale has been constructed? What arithmetic does that scale permit? The task is therefore not to reconcile two established measurement traditions. It is to introduce the measurement framework that HTA curricula across the board have omitted,

Reconstruction requires two fundamental commitments.

The first requirement is to begin with defined attributes. Before numbers are assigned or arithmetic contemplated, the attribute that is claimed to change under therapy must be specified. Is it manifest or latent? Is it unidimensional? What observations establish its possession? What empirical relationships must its numerical representation preserve? Measurement begins with the attribute, not with a questionnaire, preference exercise, algorithm, model or composite score. A multidimensional description of a health state cannot simply be declared a measure of a single attribute because an algorithm produces a number.

The second requirement is to recognize there are only two measurement pathways for therapy-impact claims. For a manifest attribute, the requirement is a linear ratio measure with an invariant unit and non-arbitrary zero, permitting the ratio arithmetic required for proportional and multiplicative claims. For a latent attribute, the appropriate pathway is Rasch measurement, where persons and items are conjointly located on a unidimensional continuum, with the requirements of invariance, specific objectivity, ordered thresholds, appropriate fit and the other measurement properties necessary to support a defensible quantitative claim. These are fundamentally different measurement problems and must not be collapsed into the legacy practice of assigning numbers and assuming that subsequent arithmetic legitimizes them.

The implications for curriculum design are substantial. Utilities, QALYs and cost-per-QALY claims cannot remain the organizing framework and then be supplemented by a short introductory discussion of measurement. That would preserve the inversion. The sequence itself must be reversed: Attribute → Measurement → Scale properties → Admissible transformations → Admissible arithmetic → Therapy-impact claim → Empirical evaluation

Students should therefore learn why scale type determines permissible arithmetic, why dimensional homogeneity and commensurability matter, why manifest and latent attributes require different measurement pathways, and why measurement claims must be capable of empirical evaluation, replication and falsification.

This is not an argument for abandoning quantitative HTA. It is an argument for making HTA genuinely quantitative by requiring its numbers to represent lawful measures before arithmetic is performed upon them. Nor does reconstruction require waiting for another generation of methodological debate. The principles of representational measurement are already available. The task is to reorganize HTA around them.

The transition can therefore begin immediately. Existing curricula should be reconsidered from their foundations rather than incrementally amended. The central educational question is no longer how students can become more proficient in manipulating the numerical constructs inherited from legacy HTA. It is whether they can identify an attribute, construct or select a lawful measure of that attribute, determine the arithmetic permitted by its scale properties, and formulate a therapy-impact claim that can survive empirical challenge. Measurement must precede arithmetic. That principle provides the starting point for reconstructing health technology assessment without the baggage of legacy HTA.

8. FACILITATING THE TRANSITION

Reconstructing HTA does not require universities, research centers or other organizations to develop a measurement-first curriculum independently. A structured program has already been developed by Maimon Research specifically to facilitate the transition from legacy HTA to an assessment framework grounded in representational measurement: *Reconstructing HTA Programs: Measurement before Arithmetic* ⁴. The program is designed not as an additional measurement module to be inserted into an otherwise unchanged HTA curriculum, but as a reconstruction of the sequence by which quantitative claims are created, evaluated and taught.

The program comprises nine integrated units of approximately 45,000 words and provides the foundation for a three-unit graduate course, faculty-development program or professional educational program. Its organizing principle is straightforward: measurement must precede arithmetic. The purpose is to provide faculty, researchers, practitioners and students with the conceptual and practical tools required to identify what is being measured, determine how it can lawfully be measured, establish the properties of the resulting scale and restrict subsequent arithmetic to operations permitted by those properties.

The program begins with the definition of attributes. This is deliberately prior to discussion of instruments, utilities, modelling or economic evaluation. A quantitative therapy-impact claim must

identify the attribute whose magnitude is claimed to change. The program then develops the requirements of representational measurement, including order, additivity, invariance, solvability, cancellation and the other properties required to justify the representation of empirical relationships numerically. Particular attention is given to scales of measurement, admissible transformations and admissible arithmetic. Students are required to understand that assigning numbers does not itself create measurement and that the permissible mathematical operations depend upon the properties of the scale that has actually been constructed.

This foundation leads directly to dimensional homogeneity and commensurability. These are treated not as peripheral mathematical considerations but as requirements governing the construction and interpretation of quantitative claims. The distinction between a dimensionless numerical result and a lawful empirical proportion is emphasized, together with the requirement that ratios intended to express proportions must be constructed from commensurable measures.

A central component of the program is the distinction between manifest and latent attributes. Manifest attributes capable of direct quantitative representation require linear ratio measurement, with an invariant unit and non-arbitrary zero. Latent attributes present a fundamentally different measurement problem. Their assessment requires a measurement model capable of constructing a unidimensional continuum rather than simply summing responses to questionnaire items. The program therefore develops the Rasch measurement framework, including person-item conjoint measurement, specific objectivity, invariance, ordered thresholds, item fit, differential item functioning and the requirements for establishing a defensible latent measurement scale.

These two pathways converge on the construction of therapy-impact claims. For manifest attributes, impact must be expressed through lawful linear ratio measurement. For latent attributes, impact must be established through Rasch measurement. The program also addresses attribute possession, an issue largely overlooked in conventional HTA: before change in an attribute can be claimed, there must be a defensible basis for establishing that the individual possesses the attribute being measured. Therapy-impact claims are then considered in terms of empirical evaluation, uncertainty, replication and falsification.

The objective is therefore not simply to demonstrate why utilities, QALYs and related legacy constructs fail. It is to provide a positive replacement framework for quantitative HTA. Students and faculty move from identifying an attribute to selecting the appropriate measurement pathway, establishing scale properties, determining admissible arithmetic and constructing claims that can be evaluated against empirical evidence. The program replaces procedural competence in inherited techniques with measurement competence.

Maimon Research can support this transition beyond provision of the nine-unit program itself. Support can include online question-and-answer sessions, seminars, faculty discussions, workshops, small-group teaching, review of existing curricula, identification of measurement and curriculum inversions, and assistance in restructuring courses around representational measurement. Institutions can therefore determine the extent and pace of their transition while drawing upon continuing support as faculty become familiar with a body of measurement knowledge that may not have formed part of their own graduate education.

This matters because the transition should not be delayed by the argument that faculty first require years to construct an alternative curriculum. The intellectual foundations are established, the educational framework has been developed, and mechanisms for continuing academic support are available. A modest program fee supports maintenance and updating of the materials together with continuing educational and institutional support.

The choice confronting HTA education is therefore no longer between retaining legacy HTA and waiting indefinitely for some future alternative. The alternative exists now. The task is to move from a curriculum in which arithmetic precedes measurement to one in which every quantitative claim begins with a defined attribute, proceeds through a lawful measurement pathway and employs only the arithmetic warranted by the resulting scale. Maimon Research provides a structured framework and continuing support for institutions prepared to make that transition.

CONCLUSION

There is no longer a choice between legacy HTA and a measurement-first reconstruction of health technology assessment. They are not competing paradigms between which faculty, students, agencies or practitioners may reasonably select according to methodological preference. The reason is straightforward: legacy HTA is a subject without measurement. Its central quantitative framework permits arithmetic to precede the demonstration of the measurement properties required to make that arithmetic meaningful. The result is a legacy of false QALY-based quantitative claims.

The QALY provides the definitive proof. A preference or utility value cannot become a proportional adjustment factor for time merely because it is represented by a number between zero and one. The value must first be established as a lawful ratio measure of a defined attribute. Even if that requirement could be satisfied, a further measurement relationship would have to demonstrate that a proportion in preference space maps to the corresponding proportion in time space. The conventional QALY establishes neither. The failure is not repairable by improved elicitation techniques, alternative value sets, more sophisticated statistical estimation or more elaborate modelling. It is inherent in the construction itself.

Nor does the problem begin with the QALY. The QALY is the tip of the iceberg. The failure begins upstream with health-state valuation, where numerical assignment has been allowed to substitute for measurement, and it propagates through preference scores, utilities, QALYs, incremental QALYs, cost-per-QALY claims, decision models and QALY-dependent reference cases. Forty years of increasingly sophisticated arithmetic cannot compensate for the absence of the logically prior requirement for lawful measurement.

This has immediate implications for education. A legacy HTA curriculum cannot reasonably be presented to students as an adequate scientific foundation for a professional career in HTA and health economics once its measurement failure has been demonstrated. Students should be expected to ask what attribute is being measured, whether it is unidimensional, what scale has been constructed, what transformations are admissible and what arithmetic that scale permits. They should expect quantitative claims to satisfy the same fundamental measurement requirements demanded elsewhere in quantitative science.

Change is therefore mandated by the requirements of measurement themselves. The issue is no longer whether HTA should change. The question is how universities, faculty and HTA research centers will respond.

Will faculty reconsider curricula that they themselves inherited and have taught for many years? Will universities recognize that procedural competence in utilities, QALYs, ICERs and decision models cannot substitute for measurement competence? Will HTA centers reconstruct their research and educational programs around defined attributes and lawful measurement? Or will the problem be placed in the too-hard basket, preserving familiar courses and methods in the hope that the underlying measurement failure will remain outside professional attention?

Students may ultimately become the catalyst. Once they understand that measurement must precede arithmetic, they can reasonably ask why they should enroll in, or pay for, a program that continues to teach quantitative procedures without first establishing that the quantities being manipulated are measures. They can demand a curriculum that prepares them for a scientific rather than merely procedural approach to therapy assessment.

The alternative is already available. HTA can be reconstructed around defined attributes and two lawful measurement pathways for therapy-impact claims: linear ratio measurement for manifest attributes and Rasch measurement for latent attributes. Scale properties determine admissible arithmetic; dimensional homogeneity and commensurability constrain quantitative claims; and those claims must be open to empirical evaluation, replication and falsification.

The responsibility now passes to universities, research centers, faculty and students. The measurement question can no longer be kept outside the intellectual boundaries of HTA. Legacy HTA cannot be defended by its longevity, institutional acceptance or global reach. Forty years of acceptance is not a measurement warrant. Once measurement precedes arithmetic, the legacy framework disappears. The task now is to replace it. There is no alternative.

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 more than 20 years teaching at 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 mosquitoes and the swimming pool.

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 of 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 Langley P. Maimon Research: *Reconstructing HTA Programs: Measurement before Arithmetic*
<https://maimonresearch.com/hta-reconstruction-program-and-fees/>