MAIMON RESEARCH LLC

ARTIFICIAL INTELLIGENCE LARGE LANGUAGE MODEL INTERROGATION

REPRESENTATIONAL MEASUREMENT AND HEALTH TECHNOLGY ASSESSMENT

NEW ZEALAND: LEGACY HTA - FORTY YEARS OF NUMERICAL STORYTELLING

The Absence of Representational Measurement in Teaching and Research

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

LOGIT WORKING PAPER No 8241 AUGUST 2026

www.maimonresearch.com

Tucson AZ

ABSTRACT

New Zealand health technology assessment (HTA) has developed within an international analytical framework centered on economic evaluation, utility measurement, quality-adjusted life years (QALYs), incremental cost-effectiveness ratios (ICERs) and decision modelling, with PHARMAC providing the principal national institutional influence on pharmaceutical economic evaluation. Despite decades of application, the defining quantitative operations of this framework have not, as far as can be established, been formally demonstrated to satisfy the requirements of representational measurement. This paper addresses that omission by applying a strict measurement sequence: defined unidimensional attribute → axiomatic requirements → measurement scale → admissible arithmetic → quantitative claim.

Ratio measurement provides the decisive test because multiplication, division and proportional interpretation require appropriate ratio properties. The conventional QALY, QALY = U × T, requires utility U *to function as a dimensionless proportional adjustment to ratio-scaled time* T. *Conventional multiattribute utility scores fail this requirement: they do not represent a demonstrated unidimensional ratio attribute, possess no established ratio-scaled numerator and denominator whose units cancel, and may take negative values incompatible with interpretation as a conventional proportion. Rasch measurement provides legitimate measurement of defined latent attributes but cannot rescue the QALY: logit differences yield invariant ratios of odds, not linear proportions of attribute possession capable of adjusting time. The two legitimate measurement pathways therefore close both routes to the QALY.*

The failure of the QALY removes ΔQALY as a defensible denominator of the conventional ICER. The composite cost numerator presents an independent problem because price-weighted aggregation of heterogeneous resource attributes does not create a unidimensional measure of resource utilization. Consequently, probabilistic modelling, sensitivity analysis and increasingly sophisticated reference cases cannot repair measurement failures that precede modelling.

For New Zealand, the implication is categorical: the appropriate response is not refinement of legacy HTA but its rejection and replacement. PHARMAC practice, international convention and academic acceptance cannot substitute for measurement proof. Universities have a corresponding scientific responsibility and duty of care to students: once these foundations have been challenged, legacy methods should not be presented as settled quantitative science without providing students the measurement principles required to evaluate them. The required transition is to attribute-specific, lawful measurement and therapy-impact claims capable of empirical evaluation, replication and falsification.

INTRODUCTION

New Zealand provides an especially important setting in which to undertake a formal reassessment of the scientific foundations of health technology assessment (HTA). Over several decades, New Zealand HTA has developed within the same international analytical framework centered on economic evaluation, utility measurement, quality-adjusted life years (QALYs), incremental cost-effectiveness ratios (ICERs) and decision modelling. Within the New Zealand health system, this framework has been given particular institutional force through PHARMAC and its approach to pharmaceutical economic evaluation and resource-allocation decision making. Although New Zealand has a substantially smaller HTA and health-economics research infrastructure than Australia, the underlying scientific question is identical: have the quantitative operations upon which this framework depends ever been demonstrated to satisfy the requirements of representational measurement [1] [2] ?

The New Zealand experience is therefore more than an example of the international diffusion of legacy HTA. It represents a long-term institutional commitment to an inherited analytical framework. PHARMAC has exercised a major influence over the form in which pharmaceutical economic evidence is generated and presented in New Zealand. Researchers, consultants and manufacturers engaging with the reimbursement process have necessarily had to understand and respond to that framework. University researchers working in health economics, outcomes research, public health and related areas have operated within the same broader international methodological tradition. As far as can be established from publicly available research, curricula and program content, there has been no corresponding New Zealand tradition in which the defining quantitative operations of legacy HTA were systematically challenged against the axioms of representational measurement.

This history matters because institutional longevity cannot substitute for scientific validity. A methodology does not acquire measurement legitimacy through repeated application, government adoption, academic acceptance or the accumulation of an extensive international literature. Nor does the fact that a methodology has become embedded in the decision-making processes of a national agency establish the scientific admissibility of its underlying arithmetic. The logically prior question remains whether the quantities being manipulated are measures of defined attributes, whether their empirical structures support the scales attributed to them and whether those scales permit the multiplication, division, subtraction and aggregation subsequently performed.

The time has therefore come for a different form of assessment. The New Zealand debate should move beyond criticism of particular modelling assumptions, utility instruments, parameter estimates, uncertainty analyses or individual funding decisions. What is required is a formal series of proofs addressing the measurement foundations of legacy HTA itself. The purpose is not to demonstrate that particular PHARMAC analyses or individual economic evaluations have been poorly executed. The question is more fundamental: can the analytical framework, even when executed exactly as intended, satisfy the requirements of representational measurement?

This paper initiates that assessment. It formally assesses the arithmetic operations defining legacy HTA against the scales of measurement and the required axioms of representational measurement. After decades in which the framework has been applied and institutionalized without an equivalent formal assessment of these underlying requirements, the question for New Zealand is elementary but decisive: can the analytical framework of legacy HTA withstand formal assessment against the required axioms of representational measurement and the scales of measurement?

This question takes precedence over the familiarity, institutional acceptance or mathematical sophistication of the methods involved. Utilities, QALYs, ICERs and economic modelling have become familiar components of health-economic evaluation. PHARMAC acceptance cannot, however, establish measurement legitimacy, any more than international or academic acceptance can. The relevant questions are whether the numbers being manipulated represent defined attributes, whether those attributes possess the empirical structure required for the claimed scales, and whether multiplication, division, subtraction and aggregation are admissible operations for those scales.

These are not standards devised retrospectively to challenge New Zealand HTA. The principles of scale theory and representational measurement were established before the QALY and cost-per-QALY framework became institutionalized internationally. The problem is therefore not that New Zealand HTA has failed to satisfy some newly invented scientific standard. It is that the inherited framework appears to have developed and been adopted without formally establishing its compliance with measurement requirements that were already available.

Quantitative science begins with an uncompromising requirement: before numbers can support a scientific claim, there must be a defined attribute capable of lawful measurement. Measurement does not begin with a questionnaire, a multiattribute health-state description, a score, a utility algorithm or an economic model. Nor can an attribute simply be nominated and assigned numbers. Its empirical structure must satisfy the requirements of representational measurement for the scale being claimed. The required sequence is:defined unidimensional attribute → axiomatic requirements → measurement scale → admissible arithmetic → quantitative claim

Legacy HTA reverses this sequence. Arithmetic operations and modelling requirements determine what numbers are required, while the prior measurement status of those numbers is assumed or left unexamined. This is measurement inversion.

The present paper applies the correct sequence to the analytical framework that has become established in New Zealand. It distinguishes two routes to quantitative measurement: linear ratio measurement for appropriate manifest attributes and Rasch logit measurement for latent attributes. It then asks whether the central constructs of legacy HTA satisfy the measurement requirements necessary for the arithmetic performed upon them.

The decisive test is the QALY because it connects multiattribute health-state modelling directly to conventional cost-effectiveness analysis:

QALY = U × T

where *U* is the utility score generated from a weighted multiattribute health-state description and *T* is the duration of time associated with that state. Utility is required to operate as a proportional adjustment to ratio-scaled time. The resulting QALYs are aggregated and compared to generate ΔQALY, which becomes the denominator of the conventional incremental cost-effectiveness ratio:

ICER = ΔC / ΔQALY

Each step makes specific measurement demands. If *U* is to adjust *T* proportionately, it must possess the required ratio properties and provide the appropriate dimensionless quantity. If QALYs are to be aggregated and differenced, the resulting quantity must retain the required measurement structure. If ΔQALY is to serve as the denominator of a ratio, that denominator must itself be a legitimate quantitative measure. These requirements cannot be supplied retrospectively by the sophistication of the economic model into which the numbers are entered.

The importance of undertaking this assessment in New Zealand is reinforced by the preceding interrogations of New Zealand HTA knowledge bases, university programs and publicly represented program content. These have identified the same fundamental inversion observed internationally: extensive representation of the concepts and techniques of contemporary HTA accompanied by the absence or near-absence of the representational-measurement principles required to establish their quantitative legitimacy. The resulting pattern is one of measurement inversion, curriculum inversion and program-content inversion: arithmetic precedes measurement; established HTA methods precede instruction in the scientific requirements that determine whether those methods are admissible; and the publicly represented intellectual infrastructure reproduces legacy HTA without providing comparable representation of the measurement standards against which it should first be judged.

The New Zealand institutional setting makes this especially important. PHARMAC occupies a distinctive position in the national medicines environment. A comparatively small health-economics and HTA community operates alongside a powerful national decision-making institution whose methodological expectations inevitably influence the form in which economic evidence is produced and discussed. Universities train students who may subsequently work within this environment; researchers contribute to the health-economic literature; consultants and manufacturers respond to institutional requirements; and successive generations encounter an established methodological framework whose legitimacy can easily appear to have been settled before they enter the field.

This should not be interpreted as evidence of deliberate scientific neglect by PHARMAC or New Zealand academics. A more plausible explanation is historical and epistemic. New Zealand inherited an international HTA paradigm in which utilities, QALYs and cost-effectiveness modelling were already accepted methodological objects. Researchers were educated within that paradigm and subsequently contributed to its continuation. What was inherited was reproduced.

That explanation, however, has a limited life. Once the measurement problem has been explicitly identified, historical inheritance may explain its origin; it cannot justify its continued transmission.

The next stage must therefore consist of explicit proofs. The question is no longer whether representational measurement deserves greater recognition within New Zealand HTA education or whether measurement theory might usefully be added to existing health-economics curricula. The issue is whether the defining operations of the inherited analytical framework survive formal measurement scrutiny at all.

The first proof concerns multiplication. Can the utility score legitimately serve as the proportional quantity required for multiplication by ratio-scaled time? The second concerns dimensional coherence. Does the QALY operation produce a quantitatively meaningful quantity rather than merely a numerical product? The third concerns aggregation and subtraction. If constituent QALYs lack the required measurement properties, can their aggregation and subsequent differencing create a legitimate ΔQALY? The fourth concerns division. Can ΔQALY serve as the denominator of a scientifically meaningful cost-effectiveness ratio? The fifth concerns the numerator itself. Does ΔC represent a defined ratio-scaled attribute capable of supporting the arithmetic required of it? The sixth concerns economic modelling as a whole. Can probabilistic simulation, sensitivity analysis and uncertainty quantification rescue outputs generated from inputs and operations whose measurement admissibility has not first been established?

The conclusions developed in this paper are categorical. Conventional utility cannot provide the dimensionless ratio quantity required to adjust time proportionately. Rasch logit measurement, although providing a legitimate measurement pathway for latent attributes, cannot rescue the QALY because a latent attribute measured through the Rasch model cannot simply be converted into the linear proportional multiplier required by the QALY construction. The QALY therefore cannot be repaired by substituting better psychometrics for conventional utility measurement.

Once the QALY fails, ΔQALY fails with it, removing the denominator of the conventional cost-per-QALY claim. The composite cost numerator presents an independent measurement problem. The conventional ICER therefore cannot acquire scientific legitimacy merely because its numerator and denominator can be represented numerically and divided by one another.

For New Zealand, the consequence is potentially profound. A national HTA environment that has operated for decades within an internationally accepted analytical paradigm must now confront a question that institutional acceptance cannot answer: were the measurement foundations of that paradigm ever scientifically admissible in the first place? If the proofs presented here stand, the appropriate response is not further refinement of the legacy framework. It is its rejection and replacement by an approach in which measurement precedes arithmetic.

1. SCIENCE BEGINS WITH THE ATTRIBUTE

Quantitative analysis does not begin with numbers. It begins with a defined attribute: the particular property whose magnitude or possession is to be represented. Survival duration, hospital days, pain, fatigue and physical function are examples. Before numbers are assigned, the investigator must specify exactly what is being measured. An attribute cannot simply be nominated and numbers attached to it. Its empirical structure must support the proposed numerical representation.

The first requirement is unidimensionality. Measurement must concern one attribute at a time. A ruler can represent differences in length because length is the attribute being measured; a Rasch ruler can represent differences in possession of physical function only if the items conform to a single defined physical-function construct. If several distinct attributes are combined, the resulting number does not become a measure merely because an algorithm produces a single numerical output. One number is not evidence of one dimension.

This point is fundamental to the subsequent critique of legacy HTA. A multiattribute health-state description deliberately combines dimensions such as mobility, self-care, usual activities, pain/discomfort and anxiety/depression. These are different attributes. Applying weights to their levels and combinations may generate a composite utility number, but weighting cannot transform multiple attributes into one unidimensional attribute. The measurement failure therefore begins before the scale properties of the utility number are even considered.

Once a unidimensional attribute has been established, the investigator must determine whether its empirical structure satisfies the axiomatic requirements of the proposed measurement representation. These include, as appropriate, order, additivity, solvability, cancellation and the Archimedean condition. Only after these requirements have been addressed does it become meaningful to ask what scale of measurement has been achieved and, consequently, what arithmetic operations are admissible.

Scales of measurement

Scales of measurement are important because numbers do not acquire identical mathematical properties merely by being numbers. Nominal, ordinal, interval and ratio scales represent different empirical structures and permit different transformations. An ordinal scale establishes order but not equal quantitative differences. An interval scale supports meaningful differences but permits an arbitrary origin. A ratio scale has the additional structure required for invariant ratio comparisons.

The permissible transformations identify what remains quantitatively meaningful. If an empirical conclusion changes merely because one permissible numerical representation has been replaced by another, that conclusion cannot be an invariant measurement claim. Scale type therefore determines the arithmetic that can legitimately be performed. Mathematical software may multiply or divide any numbers supplied to it; measurement science determines whether those operations have empirical meaning.

The Axioms of Representational Measurement

The classification of a numerical representation as a measurement scale cannot rest simply upon the appearance of the numbers assigned. Representational measurement requires a correspondence between the empirical relationships exhibited by the defined unidimensional attribute and the numerical relationships used to represent them. The numerical system must preserve the relevant structure of the empirical system. This is why the axioms of representational measurement are fundamental: they specify the conditions that the empirical attribute must satisfy before a particular numerical representation can legitimately be claimed. Measurement is therefore not achieved by assigning numbers and subsequently deciding what those numbers will mean. The empirical structure must justify the numerical structure.

For quantitative measurement, the relevant requirements include unidimensionality, order, additivity, solvability, cancellation and the Archimedean condition, as appropriate to the measurement structure being constructed, together with invariance under the permissible transformations of the resulting scale. Unidimensionality is indispensable: the relationships being represented must concern one defined attribute, not a weighted combination of different attributes. Order requires that greater and lesser possession can be consistently represented; additivity requires an appropriate concatenation or conjoint structure; solvability requires that intermediate quantitative relationships can, in principle, be represented; cancellation ensures consistency among combinations; and the Archimedean condition excludes structures in which finite comparisons cannot be established. The resulting numerical representation must then remain invariant under the transformations permitted for that scale. These requirements are not optional statistical tests added after numbers have been generated. They are the conditions that make measurement possible in the first place.

2. RATIO MEASUREMENT

Ratio measurement is the critical threshold for any analytical framework that depends upon multiplication, division or proportional interpretation. Failure to recognize this requirement virtually guarantees quantitative failure. This is precisely the problem encountered by legacy HTA. Its defining operations require utilities to act as proportional adjustments to time, QALYs to be aggregated and differenced, and incremental quantities to be divided to produce cost-effectiveness ratios. Yet the ability to perform these operations mathematically does not establish their admissibility as measurement operations. Arithmetic cannot confer ratio properties upon numbers that did not possess them before the arithmetic began**.** An analytical framework that starts by specifying the calculations it wishes to perform and only subsequently assumes that its inputs possess the necessary measurement properties has reversed the scientific sequence. Ratio measurement must precede ratio arithmetic.

Ratio measurement is therefore not defined simply by the presence of a numerical zero, by reporting positive numbers, or by the mathematical possibility of forming a quotient. A ratio measure must represent a defined unidimensional attribute whose empirical structure satisfies the axiomatic requirements necessary for quantitative representation. These include order, so that greater possession of the attribute is empirically distinguishable from lesser possession; additivity or an appropriate concatenation structure, so that quantitative differences have an empirical interpretation; the Archimedean property, ensuring that finite comparisons can be represented within the numerical system; solvability, allowing equivalent quantitative relationships to be established; and cancellation, ensuring consistency among comparisons and combinations of attribute magnitudes. The representation must also satisfy invariance, so that conclusions about the attribute do not depend upon arbitrary features of the numerical assignment, and the admissible transformations must preserve the ratio structure. For a conventional linear ratio scale this requires a meaningful non-arbitrary zero and transformations restricted to similarity transformations, $x' = ax$, $a > 0$. Consequently, statements such as *x is twice y* remain invariant under every admissible transformation.

These requirements establish a simple but uncompromising sequence: attribute → empirical relational structure → axiomatic requirements → ratio representation → admissible transformations → admissible arithmetic. Ratio status must therefore be demonstrated before multiplication, division or proportional interpretation is attempted; it cannot be assumed because an analytical framework requires those operations. If the attribute is undefined or multidimensional, if the empirical structure required by the representation has not been demonstrated, if the zero is arbitrary, or if transformations involving an additive constant are admissible, proportional statements are not invariant and ratio arithmetic is inadmissible. This is the point at which legacy HTA was effectively guaranteed to fail**.** Once utilities and other numerical constructs were permitted to enter multiplication and division without first demonstrating the required ratio properties, everything constructed downstream including the QALY, ΔQALY and the cost-per-QALY ICER became vulnerable to foundational measurement failure. The issue is therefore not whether the subsequent mathematics is sophisticated or correctly executed. It is whether the numbers entering that mathematics were measures of the required kind in the first place.

3. MANIFEST AND LATENT ATTRIBUTES

Once the attribute has been defined, its nature determines the appropriate measurement problem. The fundamental distinction is between manifest and latent attributes.

Manifest attributes are directly observable. Survival duration, hospital days, ICU hours and treatment duration are examples. Where the empirical structure satisfies the required axioms, these attributes can support conventional linear ratio measurement. There is a meaningful zero, units may be changed proportionally, and ratios remain invariant:

Latent attributes present a fundamentally different problem. Pain, fatigue, physical function and need fulfilment cannot be directly observed or physically concatenated. What can be observed are responses to items designed to provide evidence concerning possession of the specified attribute. Assigning integers to response categories or summing those integers does not create measurement. The latent attribute must first be explicitly defined and unidimensional. A measurement system must then establish an invariant continuum on which differences in possession can be represented. Rasch measurement provides this structure by locating persons and items jointly on a common logit continuum [3]. Importantly, the observations must conform sufficiently to the requirements of the measurement model. The model does not confer measurement status automatically upon whatever responses are supplied to it.

HTA therefore has two distinct routes: (i) manifest attribute → axiomatic requirements → linear ratio measure (ii) latent attribute → axiomatic/model requirements → Rasch logit measure. This distinction has a major consequence. There is no measurement requirement that survival, pain, fatigue, mobility and physical function be collapsed into one overarching number. Each is an attribute in its own right and must retain the measurement structure appropriate to it.

If HTA is to assess therapy impact on latent attributes such as pain, fatigue, physical function or need fulfilment, it requires a measurement framework capable of establishing and quantifying attribute possession. For more than fifty years, Rasch measurement has provided the necessary and sufficient framework for constructing invariant measures of possession from ordinal response observations. There is no alternative route that satisfies the same measurement requirements simply by scoring, weighting or statistically modelling questionnaire responses. Rasch measurement begins with a defined unidimensional attribute and requires the data to conform to the measurement model before claims for possession and change in possession are accepted. For readers wishing to pursue the measurement of latent-attribute possession and its application to therapy-impact claims, the Maimon Research nine-unit program provides the necessary introduction to this framework (see Section 9 below).

4. TWO ROUTES TO RATIO MEASUREMENT

The two routes provide different forms of quantitative representation, and their ratio interpretations must not be confused.

For a manifest attribute satisfying the requirements of linear ratio measurement:

$X' = aX, a > 0$

Ratios of quantities are invariant. If:

$X_1 / X_2 = 2$

then after a permissible change of unit:

$aX_1 / aX_2 = X_1 / X_2 = 2$

This structure also explains why scale properties matter for multiplication. Suppose X and Y are appropriate ratio measures and:

$Z = XY$

Under permissible transformations:

$X' = aX$

$Y' = bY$

then:

$Z' = X'Y'$

$Z' = (aX)(bY)$

$Z' = abXY$

and since:

$Z = XY$

then:

$Z' = abZ$

The derived quantity changes only by the predictable change in units.

Contrast this with a quantity permitting an arbitrary origin:

$X' = aX + c$

Then:

$Z' = (aX + c)(bY)$

$Z' = abXY + bcY$

and therefore:

$Z' = abZ + bcY$

The additional term bcY means that the transformed product is no longer simply a rescaled version of Z. The arbitrary origin contaminates the product. The ability to perform the multiplication arithmetically does not make the resulting quantity measurement-theoretically admissible.

Rasch measurement provides a different ratio structure for a latent attribute. For the dichotomous Rasch model with questionnaire items defining the assessment instrument:

$\log[Pni / (1 - Pni)] = \theta n - \beta i$

where Pni is the probability that person *n* gives the specified response to item *i*, θn is the person's location on the latent attribute and βi is the item's location on that same continuum. The expression:

$Pni / (1 - Pni)$

is the odds of the specified response. The Rasch model therefore represents the logarithm of those odds, the log odds or logit as the difference between person and item locations.

If therapy changes a person's location on the latent continuum from θ_1 before treatment to θ_2 after treatment, the change is:

$\Delta\theta = \theta_2 - \theta_1$

Because the Rasch continuum is expressed in log odds, exponentiating this difference returns the comparison to the odds scale. If the corresponding odds before and after treatment are O_1 and O_2, then:

$\theta_1 = \log(O_1)$

$\theta_2 = \log(O_2)$

Therefore:

$\Delta\theta = \log(O_2) - \log(O_1)$

Using the logarithmic identity:

$\log(A) - \log(B) = \log(A/B)$

gives:

$\Delta\theta = \log(O_2/O_1)$

Exponentiating both sides:

$\exp(\Delta\theta) = O_2/O_1$

Hence:

$OR = \exp(\Delta\theta)$

where OR is the ratio of the odds associated with the two locations.

For example, if:

$\Delta\theta = 1$

then:

$OR = \exp(1) \approx 2.72$

Equal logit differences have equal multiplicative interpretations on the odds scale regardless of where the difference occurs.

For example:

$-3 \rightarrow -2$: $\Delta\theta = -2 - (-3) = 1 \rightarrow OR = e^1 \approx 2.72$

and:

$2 \rightarrow 3$: $\Delta\theta = 3 - 2 = 1 \rightarrow OR = e^1 \approx 2.72$

Likewise, a two-logit change anywhere on the continuum gives:

$OR = e^2 \approx 7.39$

and a half-logit change gives:

$OR = e^{0.5} \approx 1.65$

The important qualification is that this does not mean that the *probability* increases by the same amount at every location. Odds and probabilities are related nonlinearly. The same one-logit change produces the same multiplicative change in odds, but different absolute changes in probability depending on the starting location.

The odds associated with the second location are therefore approximately 2.72 times the odds associated with the first. This is the ratio interpretation supplied by the Rasch logit structure. It is crucial, however, not to misinterpret it. It is a ratio of odds, not a ratio of quantities of the latent attribute itself. A change from 1 to 2 logits does not mean that the person has twice as much physical function, need fulfilment or another latent attribute. This distinction becomes decisive when considering whether a Rasch measure could ever provide the proportional adjustment to time required for a QALY.

The ratio is therefore an invariant ratio of odds. It is not a ratio of the numerical logit coordinates themselves. A person at 2 logits does not possess twice as much of the attribute as a person at 1 logit. This distinction is decisive. Linear ratio measurement permits direct ratios of quantities; Rasch logit measurement permits invariant differences and corresponding ratios of odds. Neither structure can be changed simply because a subsequent HTA calculation requires a particular type of number.

5. DIMENSIONLESS RATIOS AND DIMENSIONAL HOMOGENEITY

The preceding mathematics establishes what happens when quantities are multiplied. The next requirement is particularly important for the QALY: what kind of number can legitimately be used to adjust ratio-scaled time proportionally?

Consider a patient who spends 18 hours of a 24-hour observation period satisfying a precisely defined condition A. The proportion of time in A is:

D = time in A / total time

so:

D = 18 hours / 24 hours = 0.75

Both numerator and denominator measure the same attribute time on the same ratio scale. The units cancel:

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

D is therefore dimensionless. Its value is also invariant to the unit chosen. Expressing the same quantities in minutes gives:

D = 1080 minutes / 1440 minutes = 0.75

More generally, if:

$D = X / Y$

and X and Y are commensurable ratio measures of the same attribute, then under a permissible change of unit:

$X' = aX$

$Y' = aY$

and:

$D' = X' / Y'$

$D' = aX / aY$

$D' = X / Y = D$

The dimensionless ratio can therefore legitimately rescale time:

$T = D \times T^*$

with:

$[T] = [1][T] = [T]^*$

This is dimensional homogeneity. The calculation begins with time and ends with time because the multiplier is a genuine dimensionless proportion. This gives us a simple but demanding test. If a number such as 0.75 is to adjust time proportionally, we must be able to ask: 0.75 of what?

What is its numerator? What is its denominator? What single attribute do they measure? Are both ratio measures? What units cancel? A decimal between zero and one is not automatically a proportion. Its proportional meaning must arise from its measurement structure. This requirement provides the direct test for conventional utility.

6. THE QALY FAILS LINEAR RATIO MEASUREMENT

The conventional QALY begins not with a unidimensional ratio attribute but with a multiattribute health-state description. A utility algorithm applies weights to the dimensions and levels defining that description, conventionally anchoring full health at unity and applying decrements to generate the utility score. In simplified form: $U = 1 - \Sigma d_i$

The resulting U is therefore a numerical representation of a weighted multiattribute health-state description. It is not a demonstrated measure of a single unidimensional attribute satisfying the axiomatic requirements for linear ratio measurement.

The decrement structure may also generate: $U < 0$ with negative values conventionally representing health states worse than dead. Whatever preference interpretation is given to such numbers, this is not the structure of a conventional proportion constructed from a positive linear ratio attribute.

Yet the next step treats U precisely as a proportional adjustment:

$$QALY = U \times T$$

If:

$$U = 0.75$$

then one year is transformed into:

0.75 × 1 year = 0.75 QALYs

The critical question is now unavoidable:

What is U = 0.75 a proportion of?

A lawful dimensionless proportion would require:

$$U = X / Y$$

where X and Y are commensurable ratio measures of the same defined attribute. Then:

$$[U] = [A]/[A] = 1$$

But no such construction exists for conventional utility. There is no demonstrated ratio-scaled numerator, denominator or single attribute whose units cancel. U = 0.75 has not been demonstrated to mean 75% of anything.

The possibility of negative utilities makes the distinction especially transparent. If:

$$D = X / Y$$

with:

$$0 \leq X \leq Y$$

then:

$$0 \leq D \leq 1$$

A negative utility cannot represent a negative proportion of a positive ratio-scaled attribute. The utility number and a dimensionless ratio are different mathematical objects.

The QALY construction therefore depends upon three successive transitions, none of which has the required measurement foundation. The first is the movement from a multiattribute health-state description to a weighted utility score. Combining and weighting different health attributes may produce a single numerical value, but it does not establish a single unidimensional attribute possessing the empirical and axiomatic properties required for linear ratio measurement. The second transition occurs when this utility score is reinterpreted as a proportional adjustment to time. This requires the utility to be a dimensionless ratio constructed from commensurable ratio measures of the same attribute, yet no such ratio structure has been demonstrated. The third transition is the multiplication $U \times T$ and the declaration that the result is a QALY. Multiplication cannot supply the measurement properties missing from the utility input, nor can it create a new linear ratio attribute representing time adjusted for health-state experience. At each stage, the arithmetic proceeds without the measurement structure required to justify it.

Multiplication by time cannot manufacture the missing measurement properties. Under the requirements of linear ratio measurement and dimensional homogeneity, the conventional QALY cannot be constructed.

7. THE QALY FAILS RASCH LOGIT RATIO MEASUREMENT

The failure of conventional utility does not mean that health-related latent attributes cannot be measured. Rasch measurement provides a route to quantitative claims for defined unidimensional attributes such as pain, fatigue, physical function and need fulfilment. But Rasch measurement does not rescue the QALY. It closes the alternative route to its construction.

For the dichotomous Rasch model:

$$\log[P_{ni} / (1 - P_{ni})] = \theta_n - \beta_i$$

where θ_n represents person location and β_i item location on the latent continuum.

If therapy changes a person's location from θ_1 to θ_2:

$$\Delta\theta = \theta_2 - \theta_1$$

and the corresponding ratio interpretation is:

$$OR = \exp(\Delta\theta)$$

Suppose:

$$\theta_1 = 1 \text{ logit}$$

and:

$\theta_2 = 2$ logits

It would be incorrect to calculate:

$\theta_2 / \theta_1 = 2$

and conclude that the second location represents twice as much attribute possession. The zero of the logit continuum is not an absolute absence of the latent attribute. The legitimate quantitative statement is:

$\Delta\theta = 2 - 1 = 1$

and therefore:

$OR = \exp(1) \approx 2.72$

This is an invariant ratio of odds. It is not a statement that the person possesses 2.72 times as much physical function, pain relief or need fulfilment.

This distinction defeats any attempt to convert a Rasch measure into the proportional adjustment required by: $QALY = U \times T$

Rescaling a logit continuum to produce numbers between zero and one does not solve the problem. A transformed value of 0.75 does not thereby become 75% possession of the latent attribute. Numerical rescaling cannot create an absolute zero or convert a ratio of odds into a linear proportion.

Rasch measurement therefore permits rigorous claims concerning change in possession of a single latent attribute, but it cannot produce the dimensionless linear proportion required to adjust time. There is only one possible lawful route to such an adjustment. It must concern a single defined attribute already demonstrated to possess linear ratio properties, from which a genuine dimensionless ratio can be constructed: $D = X / Y$ where X and Y are commensurable measures of that same attribute. For latent attributes, Rasch measurement provides no such D. Nor can separate measures of pain, fatigue, physical function and other latent attributes be collapsed into a universal quality proportion.

The two legitimate measurement routes therefore close the argument. Linear ratio measurement rejects conventional utility because it is not a demonstrated dimensionless ratio. Rasch logit measurement rejects the conversion of latent-attribute possession into the linear proportion required to adjust time. There is no measurement-theoretic route from a multiattribute health-state description or a latent-attribute measure to the conventional QALY.

8. THE COLLAPSE OF THE REFERENCE CASE

The failure of the QALY removes the denominator of the conventional cost-effectiveness reference case. But the numerator presents a separate measurement failure. The familiar expression

ICER = $\Delta C / \Delta QALY$ appears to represent the division of one quantitative measure by another. It does not. The denominator has already failed. Conventional utility cannot provide the dimensionless ratio required to adjust time, while Rasch measurement cannot convert latent-attribute possession into such a proportion. Consequently: $QALY = U \times T$ cannot provide the required measure, and neither aggregation nor subtraction can repair it: $\Delta QALY = QALY_1 - QALY_0$

The cost numerator is constructed differently but encounters the same requirement that measurement begin with a defined attribute. Conventionally:

$$C = \Sigma(q_i \times p_i)$$

where q_i may represent hospital days, ICU hours, physician visits, doses, laboratory tests or procedures, and p_i represents the corresponding unit price.

The resource quantities are heterogeneous attributes. A hospital day is not a physician visit; an ICU hour is not a dose. Multiplying each by a price and adding the monetary results may provide an accounting total, but it does not create a unidimensional ratio measure of resource utilization. The appropriate therapy-impact claims remain attribute-specific: Δ hospital days; Δ ICU hours; Δ physician visits or Δ doses

There is a further requirement. Even if numerator and denominator were independently ratio measures, this would not automatically make their quotient a meaningful measure. The attributes must be compatible with the derived quantity being claimed. For example:

distance / time = speed

has the coherent dimensional structure:

$$[L]/[T] = [LT^{-1}]$$

Similarly:

$$X / Y$$

for commensurable measures of the same attribute produces:

$$[A]/[A] = 1$$

But mathematical divisibility does not mean that any ratio measure can be divided by any other ratio measure to create a scientifically meaningful new attribute. The ICER therefore fails at every level. ΔC is a price-weighted composite of heterogeneous resource inputs; $\Delta QALY$ is not a defensible ratio measure; and their quotient does not establish a measured attribute called cost-effectiveness. The reference case cannot be repaired by changing utility instruments, discount rates, model structures or uncertainty methods. Its defining quantitative architecture has failed before those refinements begin.

9. THE COST-EFFECTIVENESS DEBACLE AND NORMAL SCIENCE

The measurement failure exposes a deeper problem. The reference-case cost-effectiveness result is not designed primarily as an empirically evaluable therapy-impact claim. It is constructed to provide a decision variable: modelled ΔC / modelled $\Delta QALY$ = cost per QALY. Yet normal quantitative science requires claims that confront observation. They should be capable of evaluation, independent replication and falsification.

Consider a claim that a therapy reduces hospital utilization by two days over six months in a defined population compared with a specified alternative. The attribute and unit are explicit. The claim can be tested prospectively. Another investigator can attempt replication. Most importantly, observations can demonstrate that the claim is false.

The same principle applies to a properly constructed Rasch claim concerning a latent attribute. A claim for improvement in physical-function possession can specify the ruler, population, comparator, timeframe and expected change:

$$\Delta\theta = \theta_2 - \theta_1$$

with:

$$OR = \exp(\Delta\theta)$$

Again, the claim can confront evidence and fail.

The reference-case claim occupies a fundamentally different position from an ordinary empirical therapy-impact claim. Before questions of replication or falsification even arise, the model assumes that its numerous numerical inputs can legitimately be combined arithmetically. Health-state descriptions, utility scores, transition probabilities, event rates, resource quantities, prices, discount factors, survival estimates and other inputs are brought together within a common mathematical framework. Yet these inputs are not required to demonstrate the ratio-scale properties, dimensional compatibility or admissible transformations necessary for the operations performed upon them.

The reference case therefore fails before any question of empirical evaluation arises. Its inputs are combined arithmetically without first establishing the scale properties, dimensional relationships and permissible transformations required to justify those operations. The resulting QALY denominator is not a defensible ratio measure; the composite cost numerator is not a unidimensional measure of resource utilization; and dividing these two constructs does not create a measured attribute called cost-effectiveness.

The decisive question is therefore not whether ICER = ΔC / $\Delta QALY$ can subsequently be replicated or falsified. The prior question is more fundamental: does the quotient have any defensible quantitative meaning? If neither its construction nor the relationship between numerator and denominator satisfies the requirements of measurement, the answer is no. Arithmetic can always generate a numerical quotient, but numerical calculability does not establish measurement meaning. A number can be calculated to any desired number of decimal places and remain scientifically meaningless.

This makes the conventional discussion of uncertainty secondary. Deterministic sensitivity analysis can substitute alternative input values, and probabilistic sensitivity analysis can attach distributions to those inputs and generate distributions of model outputs. Neither addresses the prior measurement question. Uncertainty analysis around an inadmissible construct cannot make the construct admissible. Greater sophistication in manipulating the inputs cannot supply the measurement properties that were absent before the model began.

The cost-effectiveness claim therefore fails at a more elementary level than replication or falsification. There is no scientifically established quantity to replicate or falsify or as a decision construct. The reference case produces a number intended to guide a decision, but it has not demonstrated that this number represents a measurable attribute with a coherent quantitative interpretation.

That distinction should be made explicit. Empirical evaluation, replication and falsification are requirements for meaningful scientific claims. They cannot rescue a numerical construct that has already failed the prior requirements of measurement. The first obligation of HTA is therefore to establish the attribute and its lawful measurement. Only after that has been achieved does it become meaningful to ask whether the resulting therapy-impact claim survives empirical evaluation, replication and attempts at falsification.

The implications are difficult to overstate. This is not an inconsequential academic construct confined to methodological exercises. The cost-per-QALY reference case is intended to influence whether a medicine is admitted to a formulary, whether patients obtain access to it, what restrictions are placed upon its use and, directly or indirectly, the price at which it is made available. Decisions with substantial consequences for patients, manufacturers and healthcare systems are therefore being informed by a numerical construct whose claimed quantitative meaning cannot be established under the required standards of measurement.

The extraordinary feature of legacy HTA is not simply that such a construct was devised, but that it acquired institutional authority and has been taught and applied for more than forty years without its defining operations first being required to demonstrate their measurement legitimacy. A decision maker may, of course, make judgements about access and price. What cannot be defended is conferring scientific quantitative authority upon an imaginary cost-effectiveness measure merely because an arithmetic model can generate a number. If the number has no defensible measurement meaning, it has no legitimate claim to guide formulary admission or pricing as a quantitative measure of cost-effectiveness.

10. TRANSITION TO LAWFUL MEASUREMENT

The transition to lawful measurement requires more than demonstrating the measurement failures of contemporary HTA. It requires a replacement framework grounded in representational measurement. The Maimon Research nine-unit transformation program provides that framework [4]. Beginning with the principle that measurement must precede arithmetic, the program progresses through the definition of attributes, scales of measurement, admissible arithmetic, dimensional homogeneity, manifest and latent attributes, linear ratio measurement, Rasch measurement, attribute possession, and the construction of empirically evaluable therapy-impact claims. The purpose is not to add measurement to the existing HTA paradigm while leaving its analytical structure intact. It is to replace that structure with one in which quantitative claims must first satisfy the requirements of lawful measurement and can then be evaluated, falsified and replicated.

At the same time, the nine units provide the substantive foundation for a graduate course in lawful measurement for HTA. Comprising approximately 45,000 words of structured instructional material**,** the program provides sufficient depth to move beyond a brief introduction to measurement theory and support a systematic transition from the legacy HTA framework to one grounded in representational measurement. Its sequence can be translated directly into a graduate curriculum progressing from measurement preceding arithmetic, attributes and scales of measurement through admissible arithmetic and dimensional homogeneity to manifest and latent attributes, linear ratio measurement, Rasch measurement, attribute possession and the construction of empirically evaluable therapy-impact claims.

As far as can presently be ascertained, this is the only program available internationally designed specifically to support the transition from legacy HTA to an HTA grounded in the requirements of lawful measurement**.** That distinction is important. The program is not intended to add a measurement component to contemporary HTA while leaving utilities, QALYs and reference-case modelling undisturbed. It provides a structured pathway for reconsidering the scientific foundations of the discipline itself and for rebuilding quantitative therapy evaluation around measurement before arithmetic, ratio measurement and claims capable of empirical evaluation and falsification.

The resulting course would replace the conventional emphasis on utilities, mapping, QALYs and reference-case modelling with an alternative scientific structure: defining the target attribute, determining whether it is manifest or latent, constructing the appropriate linear or Rasch ratio measure, and using that measure to formulate empirically evaluable claims. The nine-unit program therefore provides not simply a critique of the legacy framework, but the content and organizing structure from which its educational replacement can be built.

Faculty and practitioners who enroll will have access to a 24/7 question-and-answer support service, providing continuing assistance as they work through the implications of representational measurement for teaching, research and HTA practice. This becomes particularly important where faculty are using the program as the foundation for curriculum development. Questions concerning course structure, individual topics, measurement concepts, examples, assignments and their application to existing HTA material can be addressed as they arise. Participants are therefore not expected to complete the nine units and then independently determine how an unfamiliar scientific framework should be translated into teaching or practice. Continuing support is intended to make the movement from understanding to implementation an integral part of the transition.

Additional faculty and institutional support can also be arranged on an agreed basis. This may include Zoom seminars and workshops, assistance with graduate-course and curriculum redesign, development of teaching materials, recommended references and readings, and guest lectures addressing particular aspects of representational measurement and its application to HTA. Support can be tailored to individual faculty members, courses, departments, research centers or professional groups. The nine-unit program should therefore be understood as the starting point for a managed transition to lawful measurement. It provides the replacement scientific framework, the substantive elements required to construct a graduate course around that framework, and continuing access to the expertise and educational resources required for implementation. The objective is not simply to demonstrate why contemporary HTA must change, but to provide faculty and practitioners with the foundations and practical means to build, teach and apply its replacement. There is no going back.

11. CONCLUSION

The extraordinary feature of New Zealand legacy HTA is therefore not simply that a methodological mistake may have been made. It is that an entire analytical architecture was progressively accepted and institutionalized before its measurement foundations had been formally secured. PHARMAC provided the national methodological center of gravity for pharmaceutical economic evaluation, while the academic and research environment, as far as can be ascertained from its publicly represented knowledge base, reproduced rather than fundamentally challenged the broader legacy architecture within which PHARMAC operated.

After several decades, institutionalization itself has become one of the principal barriers to reassessment. The framework is familiar. It is embedded in guidance and decision-making processes. Researchers have built careers around it. Universities teach its methods. Consultants apply it. Manufacturers respond to its requirements. Decision makers receive outputs expressed in its language. None of these facts, however, constitutes evidence that its quantitative operations are admissible. Scientific validity is not established by longevity, consensus, professional acceptance or institutional authority. It is established by whether the claims being made satisfy the requirements necessary for those claims to have scientific meaning.

The formal measurement sequence exposes the problem: no defined unidimensional ratio attribute → no legitimate dimensionless utility ratio → no lawful proportional adjustment to time → no QALY → no ΔQALY → no cost-per-QALY reference case

The independent measurement failure of the conventional composite cost numerator closes the remaining route to the claimed cost-effectiveness quantity. Mathematical manipulation can produce a numerical quotient, but mathematical manipulation cannot create measurement meaning that its components and their relationship do not possess. The consequence is therefore not reform, refinement or methodological accommodation. It is rejection**.**

If these proofs stand, New Zealand faces an uncomfortable conclusion. A framework that has occupied an important position in pharmaceutical economic evaluation and health-resource decision making for decades fails measurement requirements that logically should have been established before the framework was institutionalized. Decades of application within PHARMAC and the wider international HTA community cannot constitute a scientific defense. Neither can the volume of international economic evaluation, the reputation of New Zealand universities, the sophistication of the modelling community nor the international acceptance of substantially similar methods.

The appropriate response is not to assign retrospective blame. The historical development of New Zealand HTA occurred within an international intellectual environment in which the measurement foundations of the QALY and cost-effectiveness reference case were rarely subjected to the scrutiny applied here. Faculty members may reasonably argue that they themselves were educated within this framework and subsequently taught what they had been taught. Researchers and decision makers similarly inherited methods that were already regarded internationally as legitimate components of HTA. That may explain how the legacy was transmitted. It cannot justify its continued transmission once the measurement failure has been formally identified**.**

It is at this point that the responsibility of the university becomes central. A university is not merely a transmitter of established professional practice, nor is its scientific responsibility determined by the methodological requirements of a national reimbursement agency. Its responsibility is to subject professional practice to critical scientific examination and to distinguish between methods that are conventionally accepted and claims that have survived the standards necessary for scientific validity. Academic freedom unquestionably protects the right of faculty to defend conventional utility measurement, QALYs, ICERs and cost-effectiveness modelling. It equally requires an intellectual environment in which their foundations can be challenged. What cannot reasonably be defended is the continued presentation of a contested quantitative framework to students as settled scientific methodology while withholding or ignoring the measurement principles by which its admissibility can be assessed.

This creates what can appropriately be described as a duty of care to students. Students enter New Zealand universities on the reasonable expectation that quantitative methods presented to them as scientific methods have either satisfied the relevant scientific standards or, where those standards are contested, that the nature of the dispute will be disclosed. They should not be placed in the position of investing years in graduate education, doctoral research and professional specialization without being informed that the measurement legitimacy of the framework upon which that specialization may depend has been fundamentally challenged. The greater the potential consequences for their subsequent careers, the stronger the obligation to ensure that they are equipped to understand and evaluate that challenge for themselves.

Those consequences may be particularly important in a relatively small professional community. Current students may be preparing for careers in health economics, outcomes research, pharmaceutical evaluation or health-policy analysis in which familiarity with QALYs, cost-effectiveness modelling and PHARMAC requirements is regarded as an important professional competence. Recent graduates may already have invested heavily in acquiring those skills. Doctoral candidates may be constructing dissertations around methods whose measurement foundations are now disputed. Early-career academics may be building publication programs around the same framework. Future generations may enter programs designed around methodological assumptions inherited from their predecessors. If the formal measurement critique is correct, continuing to reproduce the legacy without disclosure does not protect these students. It increases their exposure to a framework whose scientific standing may become progressively more difficult to defend.

The responsibility extends to current generations of students as well as those who have yet to enrol. New Zealand universities cannot reasonably wait for PHARMAC, professional organizations or the international HTA community to determine collectively how they wish to respond. The scientific question precedes the institutional response. Universities possess their own responsibilities for the intellectual integrity of their curricula, the standards applied to research undertaken under their names and the quality of the scientific education provided to their students. PHARMAC may determine the economic evidence it requires for its decision-making processes; it cannot determine what a university must accept as valid measurement science.

Nor does the university discharge its responsibility simply by adding a lecture on measurement theory to an otherwise unchanged health-economics or HTA curriculum. Once students understand that measurement precedes arithmetic, that the attribute must first be defined, that unidimensionality is a prerequisite for measurement, that scale properties determine admissible transformations, that multiplication and division require appropriate ratio properties, and that dimensional homogeneity must be respected, they can immediately apply those principles to the remainder of their education. They can ask whether utility scores possess the properties required for multiplication by time, whether QALYs constitute legitimate quantities, whether ΔQALY can support division and whether the resulting cost-per-QALY ICER has the quantitative interpretation claimed for it. Measurement cannot be introduced seriously and then quarantined from the methods it calls into question.

The university's duty is therefore not to dictate the answer. It is to ensure that the question can no longer be avoided. Faculty who believe that the proofs presented here are mistaken should identify precisely where they fail. If utility is a ratio measure, its empirical and axiomatic foundations should be demonstrated. If multiplication by time is admissible, the required scale properties and dimensional relationship should be established. If the QALY is a legitimate quantity, its measurement structure should be specified. If ΔQALY supports a cost-effectiveness ratio, that operation should be defended within representational measurement. Appeal to PHARMAC practice, international convention, methodological guidance, accumulated publications or professional consensus is not a substitute for such a demonstration.

This is why the New Zealand response should be one of open scientific interrogation rather than institutional defense. The measurement critique should be made available to students, incorporated into graduate discussion and subjected to formal examination by faculty. Students should be taught sufficient representational measurement to evaluate both the legacy framework and its critique. They should be encouraged to identify errors in the proofs if errors exist. That is the normal operation of science: propositions are exposed to criticism, counterarguments are formulated, claims capable of falsification are tested and inadequate theories are rejected.

There is also a responsibility to those already committed professionally to HTA and health economics. Rejection of the legacy framework does not imply rejection of the accumulated expertise of New Zealand researchers, faculty or graduates. Skills in clinical evaluation, epidemiology, trial design, data analysis, health-services research, decision making and pharmaceutical policy remain valuable. What must change is the measurement foundation upon which quantitative therapy-impact claims are constructed. New Zealand universities therefore have an opportunity, at long last, not merely to acknowledge a failure but to lead the intellectual transition from an inherited framework to one capable of satisfying the standards of normal science.

That transition begins with the attribute rather than the model. It proceeds from a defined unidimensional attribute to the axioms required for its measurement; from those axioms to an appropriate linear ratio or Rasch logit measurement pathway; and only then to arithmetic operations admissible for the scale achieved. The endpoint is not an imaginary universal cost-effectiveness number but attribute-specific therapy-impact claims capable of empirical evaluation, replication and falsification and, through that process, contributing to the evolution of objective knowledge.

New Zealand therefore faces more than a methodological choice. Its universities face an educational and scientific responsibility. They can continue transmitting an inherited analytical legacy because PHARMAC employs it, professional practice recognizes it and previous generations accepted it. Or they can recognize that once the measurement foundations of that legacy have been formally challenged, students are entitled to know that the challenge exists and to possess the scientific tools necessary to judge it.

The duty of the university is not to protect legacy HTA from uncomfortable questions, nor to protect students from discovering that an established professional framework may be scientifically defective. Its duty is to protect the conditions under which valid scientific inquiry is possible and to ensure that students are not knowingly educated within a contested quantitative framework without being given the means to evaluate its foundations. Once the measurement problem has been exposed, another generation cannot simply be taught the same framework as though nothing has changed.

That is the responsibility now confronting New Zealand universities. The transition required is from decades of inherited and institutionalized legacy methodology to a science of therapy evaluation in which measurement precedes arithmetic, scientific validity precedes professional convention, and the educational interests of current and future students take precedence over preservation of an inherited analytical framework.

ACKNOWLEDGEMENT

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

REFERENCES

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

[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]Maimon Research. HTA Reconstruction. https://maimonresearch.com/hta-reconstruction-program-and-fees/