MAIMON RESEARCH LLC

ARTIFICIAL INTELLIGENCE LARGE LANGUAGE MODEL INTERROGATION

THE MEASUREMENT FUTURE IN HEALTH TECHNOLGY ASSESSMENT

SPAIN: 40 YEARS OF QALY MEASUREMENT FAILURE AND THE END OF THE LEGACY HTA CURRICULUM

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

LOGIT WORKING PAPER No 851 SEPTEMBER 2026

www.maimonresearch.com

Tucson AZ

ABSTRACT

Background: *Spanish health technology assessment (HTA) has developed a sophisticated analytical framework encompassing economic evaluation, health-state preferences, utilities, QALYs/AVACs, cost-effectiveness and cost-utility analysis, ICERs, decision modelling and uncertainty analysis. Yet the scientific legitimacy of these methods depends upon logically prior requirements of representational measurement and the scales of measurement. Eight Spanish Logit papers have examined measurement and curriculum inversion within this environment, while a pooled program-content assessment identified 0/17 representational-measurement topics compared with 20/20 contemporary HTA and health-economics topics.*

Objectives: *To assess whether the QALY can exist as a lawful measure, examine the consequences for Spanish legacy HTA and its university curriculum, and identify a framework for transition to lawful measurement.*

Methods: *The construction QALY = T × U was assessed against requirements for representational measurement, ratio measurement, dimensional homogeneity, attribute-specific measurement and commensurability. Particular attention was given to the distinction between lawful proportional relationships among ratio measures of duration within Time Space and numerical health-state valuations originating in Preference Space. The findings were considered alongside the Spanish program-content assessment and eight Spanish Logit papers.*

Results: *The QALY cannot exist as a lawful measure. A health-state preference value U = 0.5 does not establish that anything has been measured as one-half and cannot acquire the proportional meaning of the lawful time ratio 6/12 merely because the same numeral is produced. Numerical identity is not measurement identity. Even if lawful ratio measurement of a preference attribute were hypothetically achieved, a preference ratio would remain a relationship within Preference Space and could not become a time ratio. There was therefore no missing measurement proof capable of validating the QALY construction. No lawful QALY implies no lawful incremental QALY and no scientifically meaningful cost-per-QALY ICER. Reference-case modelling, simulation and uncertainty analysis cannot repair this upstream measurement failure.*

Conclusions: *The Spanish 0/17 versus 20/20 finding represents more than curriculum imbalance: it identifies program-content inversion in which downstream analytical methods are extensively represented while their logically prior measurement gatekeeper is apparently absent. Curriculum supplementation is insufficient. Measurement is not topic number 38**.** Spanish HTA requires reconstruction around defined attributes, linear ratio measurement for manifest attributes, Rasch measurement for latent attributes, admissible arithmetic, and empirically evaluable, replicable and falsifiable therapy-impact claims. A nine-unit reconstruction framework is available now. The institutional transition may be demanding; the scientific requirement is straightforward: measurement must precede arithmetic**.***

INTRODUCTION

Spanish health technology assessment (HTA) teaching, research and practice now face an argument that goes to the foundations of the discipline. Legacy HTA presents itself as a quantitative framework while failing to establish the measurement foundations required for its central quantitative claims: representational measurement and the scales of measurement [1 2] . Eight Spanish Logit papers have documented aspects of measurement inversion, curriculum inversion and the apparent absence of lawful measurement standards from the Spanish HTA and health-economics knowledge base. Yet the established framework of health-state preferences, utilities, QALYs, cost-per-QALY claims and reference-case economic modelling remains deeply embedded in Spanish HTA.

If Spain is to make the necessary transition to an HTA curriculum grounded in lawful measurement, a catalyst may now be available. A recent evaluation of the measurement basis of the QALY demonstrates that the problem is not merely that the QALY is imperfect, controversial or inadequately validated. The QALY has never existed as a lawful measurement construct [3] . More importantly, there was no missing measurement proof that might eventually have justified it. There was no possible proof**.** Given the nature of the preference scores employed and the quantitative role assigned to them, the procedure itself is inadmissible.

For more than 40 years, an apparent failure to recognize the requirements of representational measurement and the constraints imposed by scales of measurement allowed the QALY to occupy a central position in legacy HTA. Spain adopted and institutionalized this framework across health economics, university teaching, economic evaluation and national and regional technology assessment. The eight Spanish Logit papers provide a developing body of evidence that the measurement assumptions underlying that inheritance were not subjected to the logically prior standards required for lawful quantitative claims.

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

The contrast is particularly striking because Spain cannot plausibly be characterized as lacking methodological sophistication. Its universities, health-economics programs and national and regional HTA environment encompass economic evaluation, cost-effectiveness and cost-utility analysis, health-state preferences, QALYs/AVACs, ICERs, patient-reported outcomes, decision analysis, Markov modelling, simulation, uncertainty analysis, budget impact, reimbursement and resource-allocation decision making. The analytical apparatus is extensive. What the program-content interrogation fails to identify is its logically prior measurement gatekeeper.

This is an extraordinary achievement in the history of the quantitative sciences. HTA developed an increasingly sophisticated analytical apparatus while largely ignoring the measurement requirements governing whether its numbers represented measures and whether the interpretations

placed upon its arithmetic were empirically meaningful. These are not recently invented objections. The development of measurement theory and the formal analysis of scales established standards decades ago requiring the properties of numerical representations to be established before the quantitative interpretations placed upon them can be justified. Dimensional considerations impose additional constraints upon relationships between quantities.

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

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

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

The implications for Spain are therefore no longer confined to methodological debate. Eight Logit papers have already documented the broader measurement and curriculum problem. The 0/17 versus 20/20 program-content finding demonstrates its educational expression. The demonstration that the QALY itself cannot exist as a lawful measure now exposes the consequence at the center of the legacy framework.

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

The educational consequences are equally direct. Spanish universities cannot resolve the problem by adding a lecture on measurement theory to an otherwise unchanged HTA curriculum. Measurement is not topic number 38**.** It is the gatekeeper that determines whether the other quantitative topics survive scientific scrutiny. Once measurement precedes arithmetic, the starting point must become the definition of the attribute, determination of whether it is manifest or latent, construction of the appropriate lawful measure, identification of admissible arithmetic and formulation of therapy-impact claims capable of empirical evaluation, replication and falsification.

The eight Spanish Logit papers and the new program-content assessment therefore establish a sequence of evidence that is becoming increasingly difficult to dismiss. The first stage identified measurement inversion. The second exposed its reproduction through curriculum and program content. The QALY analysis now demonstrates the consequence of allowing that inversion to persist: a foundational quantity employed for decades could not satisfy the measurement requirements necessary for the quantitative function assigned to it.

The question facing Spain has consequently changed. It is no longer whether representational measurement deserves greater attention within the established HTA framework. The issue is whether that framework can survive once representational measurement is restored to its logically prior position.

Measurement must precede arithmetic. Legacy HTA reversed that order. The issue for Spain is now whether its universities, research centers and HTA institutions are prepared to reverse it back.

UNDERSTANDING MEASUREMENT

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

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

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

Spanish legacy HTA nevertheless proceeds as though this distinction did not matter. Preference scores are treated as utilities; utilities are multiplied by time; QALYs/AVACs are subtracted to create incremental QALYs; and incremental costs are divided by these quantities to create cost-per-QALY ICERs. These operations are then incorporated into economic evaluations, reference-case analyses and decision models intended to inform national and regional health technology assessment, reimbursement and resource-allocation decisions. The arithmetic becomes progressively more elaborate while the measurement failure at its foundation remains untouched.

The significance for Spain is reinforced by the eight Spanish Logit papers that have already examined different aspects of its HTA and health-economics knowledge base, teaching, research and program content. These investigations have repeatedly identified the same underlying problem: extensive engagement with the downstream methods of contemporary HTA without corresponding recognition of the representational-measurement requirements that should determine whether the quantities entering those methods support the arithmetic and interpretations imposed upon them.

The Spanish program-content assessment makes the contrast particularly difficult to dismiss. Across the pooled publicly available Spanish HTA and health-economics knowledge base examined, the result was 0/17 representational-measurement topics and 20/20 contemporary HTA and health-economics topics. The issue is therefore not a lack of analytical sophistication. Spain plainly possesses an extensive academic and institutional infrastructure encompassing health economics, economic evaluation, cost-effectiveness and cost-utility analysis, pharmacoeconomics, outcomes research, decision modelling, uncertainty analysis, budget impact and technology assessment. The problem is the inversion of the intellectual sequence required of a quantitative science. The downstream apparatus is extensively represented while the logically prior measurement framework required to establish its quantitative legitimacy is apparently absent.

The issue is therefore not simply whether Spanish analysts perform the accepted calculations correctly. They undoubtedly can. The prior question is whether the numbers entering those calculations are measures possessing the properties required for the quantitative claims being made. A correctly executed cost-utility model cannot establish retrospectively that its utility input possesses the measurement properties required for its assigned quantitative role. Probabilistic sensitivity analysis cannot manufacture a lawful QALY. Increasing computational sophistication cannot repair a measurement failure that occurred before the arithmetic began.

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

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

This is measurement inversion: arithmetic is performed first and the measurement legitimacy of the quantities upon which it depends is either assumed or left unexamined. The eight Spanish Logit papers, reinforced by the **0/17 versus 20/20** program-content result, indicate that this inversion is not an isolated methodological oversight. It is embedded in the publicly represented knowledge

and educational framework through which legacy HTA has been taught, researched and reproduced in Spain.

The consequences extend beyond university teaching. Spain has an established national and regional HTA environment in which economic evaluation contributes to technology assessment, reimbursement and healthcare decision making. The same measurement requirements apply irrespective of institutional setting. A quantity that fails the measurement test in a university classroom does not acquire measurement legitimacy when incorporated into an HTA submission, an economic model, a methodological guideline or a reimbursement decision.

Measurement must precede arithmetic. Once that requirement is imposed, the conventional QALY becomes the critical test that Spanish legacy HTA cannot pass. The issue is no longer whether Spain possesses a sophisticated HTA infrastructure. It plainly does. The issue is whether that infrastructure can demonstrate the measurement legitimacy of the quantitative constructs upon which its central claims depend. The eight Spanish Logit papers and the program-content assessment indicate why that question can no longer be avoided.

THE IMPOSSIBLE TRANSFORMATION

For more than four decades, thousands of teachers, researchers and analysts have accepted the proposition that the QALY can simply be defined as the product of time and a health-state preference value [4]:

$$QALY = T \times U$$

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

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

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

This is the sleight of hand at the center of the QALY. A valuation of a composite health state is transformed by definition into something it has never been demonstrated to be and, given its construction, cannot be a lawful proportional measure capable of adjusting time. Twelve months

multiplied by 0.5 certainly produces the number six. What has not been produced is a measured quantity corresponding to six months of "quality-adjusted" time.

There is consequently no mystery awaiting resolution by additional empirical research. There is no missing validation study, alternative tariff, larger preference sample or more sophisticated statistical model that can supply the measurement properties absent from the construction itself. There was no missing proof. There was no possible proof.

The QALY has therefore never existed as a meaningful measurement construct. What has existed is a numerical convention in which an arbitrary preference-based valuation is multiplied by elapsed time and the product is given the name "quality-adjusted life years." For more than 40 years, Australian HTA has accepted that convention as measurement. The question now is not how the QALY should be improved. It is why Australian universities, research centers and HTA agencies should continue to teach and apply a quantitative construction that cannot satisfy the measurement requirements necessary for its claimed interpretation.

DIMENSIONAL HOMOGENEITY

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

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

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

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

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

The utility must therefore function as a dimensionless factor. But this requirement does not establish that U is a lawful dimensionless proportional measure. That requires a measurement foundation of its own. A value of U = 0.5 must possess an empirical basis for meaning one-half if it is to reduce twelve months proportionally to six months.

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

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

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

Dimensional homogeneity therefore does not rescue the QALY. It exposes the problem**.** If the product is claimed to remain a measure in time, the multiplier must possess the legitimate dimensionless proportional meaning required to adjust time. The conventional preference-based utility does not.

Once again, the conclusion is unavoidable: The ability to multiply T by U demonstrates only that two numbers can be multiplied. It does not demonstrate that the resulting number is a measure of quality-adjusted time.

ATTRIBUTE AND MEASUREMENT SPACE

A further requirement is recognition of the attribute space within which a measure has meaning. Measurement does not produce numbers that are free to acquire whatever interpretation is subsequently convenient. A measure represents an identified attribute within a defined measurement structure. Its units, zero, ordering, ratios and admissible operations derive their meaning from that structure. A numerical value established in one attribute space cannot simply be transferred to another attribute space and assumed to retain, or acquire, the quantitative properties required there.

Consider time. Duration occupies a clearly defined Time Space. If T1 and T0 are lawful ratio measures of duration, then T1 / T0 is a comparison within that same attribute space. If T1 = 6 months and T0 = 12 months then T1 / T0 = 0.5. The value 0.5 has an unambiguous proportional interpretation within Time Space where T1 = 0.5 × T0. Six months is one-half of twelve months. The proportional adjustment is justified because both quantities are lawful ratio measures of the same attribute. The proportion has been created from measurement within Time Space.

Now consider the health-state preference score used in the QALY. It originates in an entirely different Preference Space. Respondents assign values to composite descriptions of health states, and the resulting numbers are represented conventionally on a scale bounded by values such as zero and one. Whatever interpretation may legitimately be attached to those numbers within Preference Space, they are not measures of duration.

Yet the QALY requires a preference number to perform precisely the proportional function required within Time Space. If U = 0.5 and T = 12 months, the QALY procedure asserts Tadjusted = 0.5 × 12 months = 6 months.

For this to represent a lawful proportional adjustment of time, the value 0.5 must have the meaning required by Tadjusted / T = 0.5. But that is a statement about a proportion within Time Space. The number 0.5 generated in Preference Space does not become a time proportion merely because it has the same numerical value as the ratio 6/12. This distinction is fundamental 0.5 in Preference Space ≠ 0.5 as a proportional measure in Time Space. Numerical identity is not measurement identity.

The problem is even more decisive because the conventional health-state preference value is not itself a lawful ratio measure of a defined unidimensional preference attribute. Its value of 0.5 therefore does not establish “one-half” even within Preference Space. The QALY then takes this arbitrary preference-based number and assigns it the additional role of representing one-half of elapsed time.

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

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

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

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

COMMENSURABILITY

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

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

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

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

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

THE QALY CANNOT EXIST AS A LAWFUL MEASURE

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

This eliminates the final possible defense of the QALY. Giving preference values lawful ratio properties would not solve the problem. Ratio measurement establishes proportional meaning within the attribute space being measured. It does not authorize that proportion to be reassigned as the proportional adjustment of a different attribute. The QALY nevertheless requires exactly this unsupported substitution. Given QALY = T × U a value U = 0.5 is used to assert Tadjusted = 0.5

× T and therefore Tadjusted / T = 0.5. But the proportional relationship required here is a relationship in Time Space. A ratio established in Preference Space cannot, merely by numerical substitution, become a ratio in Time Space.

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

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

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

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

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

SPAIN: FAILED CURRICULUM CONTENT

The QALY measurement failure identified in the preceding sections cannot be separated from what Spanish universities and HTA programs teach. The program-content evidence has now been assembled for the pooled publicly available Spanish university, HTA and health-economics research and teaching environment. This assessment forms part of a wider body of eight Spanish Logit papers examining measurement inversion, curriculum inversion and the knowledge

foundations of legacy HTA. The program-content result could hardly have been more striking: 0/17 for representational-measurement content and 20/20 for contemporary HTA and health-economics content**.**

The assessment examined the collective publicly available teaching, training, research and methodological content through which the Spanish HTA and health-economics knowledge base is expressed and reproduced. Importantly, the institutions and programs were pooled. A topic did not have to appear across all Spanish universities or research environments to qualify as represented. Substantive evidence from any relevant part of the pooled knowledge base was sufficient for the topic to be recorded as present. The test therefore gave the Spanish academic HTA and health-economics community every reasonable opportunity to demonstrate that representational measurement formed part of its collective knowledge base.

It did not.

The first assessment comprised 17 topics representing the foundations required for lawful measurement. These included measurement preceding arithmetic, attribute definition, scales of measurement, admissible transformations, ratio measurement, dimensional homogeneity, the distinction between manifest and latent attributes, linear ratio measurement, Rasch measurement, person-item conjoint measurement, specific objectivity, invariance and attribute possession.

The result was: Representational measurement: 0/17

The second assessment comprised 20 topics characterizing contemporary HTA and health economics. These included health economics, HTA, economic evaluation, cost-effectiveness analysis, cost-utility analysis, pharmacoeconomics, utilities and health-state preferences, QALYs/AVACs, ICERs, patient-reported outcomes, decision analysis, decision trees, Markov models, simulation, uncertainty analysis, value-of-information concepts, budget-impact analysis, resource allocation, value assessment and reimbursement decision making.

The result was: Contemporary HTA and health economics: 20/20

This is not evidence of an unsophisticated Spanish HTA community. It demonstrates something considerably more troubling: comprehensive representation of the downstream analytical apparatus accompanied by an apparent absence of the logically prior measurement framework required to determine whether that apparatus has quantitative legitimacy. Spain has developed a substantial academic and institutional environment encompassing health economics, economic evaluation, pharmacoeconomics, outcomes research, decision modelling and national and regional health technology assessment. The analytical sophistication is not in question. Its measurement foundation is.

The result provides an unusually clear example of curriculum inversion. Students can be introduced to utilities before establishing what constitutes a lawful measure; to QALYs/AVACs before establishing whether the preference multiplier possesses the proportional properties required of it; to ICERs before establishing whether the denominator represents a legitimate quantity; and to increasingly sophisticated decision models before being taught that arithmetic

operations acquire empirical meaning only through the properties of the measures upon which they operate.

The 0/17 result also helps explain the otherwise extraordinary survival of the QALY in Spanish HTA. If students and faculty are not introduced to representational measurement, ratio measurement, dimensional homogeneity, commensurability and attribute-specific measurement requirements, they lack the conceptual framework necessary to identify the fundamental error in the QALY construction. The QALY survived not because Spanish HTA demonstrated its measurement legitimacy, but within an academic and methodological environment in which the knowledge required to challenge that legitimacy was apparently absent from the publicly documented program content. There was no recognition of the impossibility of using a preference value as though it were the proportional adjustment required to create quality-adjusted time.

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

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

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

The significance of the 0/17 result has therefore changed. When first presented, it documented a major omission in Spanish HTA and health-economics program content. It must now be understood as evidence of the curricular environment that allowed an impossible measurement construction to be taught and applied without its foundational error being recognized. The failure has disadvantaged not only generations of students but also generations of faculty. Many of those now teaching utilities, QALYs, ICERs and economic modelling are teaching what they themselves were taught. If representational measurement, ratio measurement, dimensional homogeneity, commensurability and the distinction between manifest and latent attributes were absent from their

own education, they were denied the intellectual tools required to interrogate the measurement legitimacy of the framework they subsequently inherited and taught.

The resulting curriculum failure is therefore self-replicating: faculty reproduce what previous faculty taught them, students inherit the same analytical framework, and institutional continuity is mistaken for scientific validation. The framework is reinforced through university teaching, methodological guidance, journals, professional organizations and the wider national and regional HTA environment. Repetition then creates the appearance of validation when what has actually been replicated is the methodology rather than the measurement proof upon which its legitimacy should depend.

This explains how measurement inversion could persist across generations without requiring incompetence or bad faith on the part of individual Spanish faculty. But the explanation has an important limit. Once the missing measurement framework has been identified and its consequences made explicit, inheritance can no longer justify reproduction. What faculty were taught explains the past; it cannot determine what they teach next.

Nor can the result now be treated simply as historical evidence. The Spanish academic and HTA community has been presented with a developing body of evidence. Eight Spanish Logit papers have examined different aspects of measurement inversion, curriculum inversion and the knowledge foundations of legacy HTA. The program-content interrogation has now demonstrated the stark 0/17 versus 20/20 inversion. The absence of representational measurement has been identified. Its consequences for utilities, QALYs, incremental QALYs, cost-per-QALY ICERs and the wider modelling framework have been made explicit. Ignorance may explain the original curriculum. It cannot justify its continuation after the omission and its consequences have been brought to the attention of those responsible for teaching and research.

This places Spanish universities and HTA teaching environments in a fundamentally different position as they approach the 2027 academic year. The question is no longer whether representational measurement would be an interesting addition to existing HTA teaching. The question is whether the existing curriculum can survive once representational measurement becomes its gatekeeper. It cannot.

The problem cannot be resolved by moving from 0/17 to 2/17 or 5/17 while retaining the existing analytical architecture. Measurement is not topic number 38**.** Representational measurement is not supplementary material to be appended to the 20/20 legacy curriculum. It determines whether those 20 topics, and the quantitative claims made within them, are admissible in the first place. Once students understand that measurement precedes arithmetic, that attributes must be identified, that manifest and latent attributes require different measurement pathways, that scale properties constrain arithmetic, and that proportional claims require appropriate measurement structures, they will inevitably apply those standards to the methods they are subsequently taught.

The contradiction will then be obvious.

Spain cannot coherently teach measurement first and legacy HTA second without requiring legacy HTA to pass the measurement standards it previously ignored. Once that test is imposed, its central

QALY-based architecture fails. QALYs cannot simply continue to be taught as measures; incremental QALYs cannot be assumed to be legitimate quantities; and cost-per-QALY ICERs cannot retain scientific meaning merely because they remain components of established economic-evaluation practice.

The 0/17 versus 20/20 finding therefore represents more than curriculum imbalance. It identifies the knowledge failure that permitted measurement inversion to become normal HTA practice. The appropriate response is not curriculum supplementation. It is curriculum reconstruction**.**

The baseline has now been established. Eight Spanish Logit papers have documented different aspects of the measurement and knowledge problem. The 0/17 versus 20/20 result provides direct evidence of program-content inversion. The QALY analysis identifies the consequence of that inversion at the center of the legacy framework. The Spanish HTA academic community can no longer treat the measurement issue as an unexplored methodological concern. The measurement failure is explicit.

The issue for 2027 is therefore not whether Spanish HTA programs will add representational measurement to legacy HTA. It is whether they will begin replacing a failed 20/20 legacy curriculum with an assessment framework in which lawful measurement determines what may legitimately be taught as quantitative science. That reconstruction begins with defined attributes, distinguishes manifest from latent measurement, employs appropriate linear ratio measurement for manifest attributes and Rasch measurement for latent attributes, establishes admissible arithmetic before quantitative claims are made, and requires those claims to be empirically evaluable, replicable and capable of falsification. Spanish universities, faculty and students should accept no less.

SPAIN: THE TRANSFORMATION TO LAWFUL MEASUREMENT

Abandonment of the legacy HTA curriculum and the QALY does not leave Spanish universities and research centers without a way forward. The required transformation has already been set out in a nine-unit program, *Reconstructing HTA Programs: Measurement before Arithmetic* [5] . Its importance is not that it offers another variant of conventional HTA education. It does the opposite. It abandons the inherited sequence and reconstructs HTA from the measurement foundations that should have preceded quantitative analysis from the outset. Utilities, QALYs and reference-case simulations have no part to play.

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

The nine-unit program develops this reconstruction progressively. It begins with measurement inversion and the consequences of allowing arithmetic to precede measurement. It then establishes

the importance of attributes, scale types, admissible transformations, dimensional homogeneity and the restrictions that measurement properties impose upon quantitative claims.

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

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

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

The new curriculum presents a much more straightforward framework for therapy-impact assessment. The starting point is the attribute about which the therapy-impact claim is to be made. If the selected attribute is manifest, such as survival time, hospital days, walking distance or treatment administrations, it must be represented by an appropriate linear ratio measure, with equal units, a non-arbitrary zero and the admissible transformations required for meaningful proportional comparisons. If the selected attribute is latent, such as pain, fatigue, physical function or need fulfilment, it requires the Rasch measurement model capable of constructing a unidimensional logit scale through person-item conjoint measurement, invariance and specific objectivity; multiplicative comparisons can then be expressed through the corresponding odds [6]. These two pathways linear ratio measurement for manifest attributes and Rasch measurement for latent attributes—provide the measurement structures required for therapy-impact claims under the standards of representational measurement.

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

This distinction is critical for Spanish universities. The transformation does not require waiting for Spain's national or regional HTA institutions, ISPOR or any other organization to authorize a new scientific standard. The standards of lawful measurement already exist. Nor is there a requirement to wait for another generation of methodological debate. Once a quantitative operation has been

shown to lack the measurement properties necessary for its claimed interpretation, institutional convention cannot preserve its scientific legitimacy.

The case for transition in Spain is particularly difficult to postpone because the problem is no longer unidentified. Eight Spanish logit papers have examined different aspects of the Spanish HTA and health-economics knowledge base, measurement framework and program content. The 0/17 versus 20/20 program-content result has established the extent of curriculum inversion, while the QALY analysis now demonstrates the consequences of that inversion for the central quantitative construct of legacy cost-utility HTA. The missing measurement framework has been identified; the consequences of its absence have been demonstrated; and a framework for replacement is available.

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

The nine-unit program demonstrates that the transition can begin immediately. It provides faculty with a coherent structure for replacing measurement inversion with measurement discipline and students with the intellectual tools required to distinguish a numerical calculation from a scientifically defensible quantitative claim. It also breaks the cycle through which faculty reproduce the framework they themselves inherited and another generation of students is trained to accept numerical convention as measurement.

For Spain, this is particularly important because the legacy framework is embedded not simply in isolated courses but across a sophisticated health-economics and HTA environment. Economic evaluation, cost-effectiveness and cost-utility analysis, QALYs/AVACs, ICERs, patient-reported outcomes, decision modelling, uncertainty analysis, budget-impact analysis and reimbursement decision making are extensively represented. The 20/20 result demonstrates the reach of this downstream analytical apparatus. The 0/17 result demonstrates the apparent absence of the measurement gatekeeper that should have preceded it.

Spanish HTA therefore faces not a choice between two competing schools of economic evaluation. There is only one scientifically defensible direction of travel: from a legacy curriculum constructed around numerical convention to an assessment framework constructed around lawful measurement.

The institutional transition may be demanding. The scientific requirement is not. After eight Spanish logit papers, a 0/17 versus 20/20 program-content assessment and the demonstration that the QALY cannot exist as a lawful measure, the issue is no longer whether the measurement failure has been identified. It is whether Spanish universities and research centers will continue teaching a legacy framework after the scientific basis for its central quantitative construct has failed, or whether they will now begin its replacement.

What faculty were taught explains the past; it cannot determine what they teach next. Measurement must precede arithmetic. The framework for making that transition is available now.

ACKNOWLEDGEMENT

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

ABOUT THE AUTHOR

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

REFERENCES

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

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

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

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

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

[6] Bond T, Yan Z, Heene M. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. New York: Routledge, 2021