

Metrological references for health care based on entropy

Leslie R Pendrill^{1,*}, Jeanette Melin¹, Stefan J Cano² and the EMPIR NeuroMET 15HLT04 consortium³

¹RI.SE Metrology, Eklandagatan 86, S-412 61, Göteborg, Sweden

²Modus Outcomes, Spirella Building, Letchworth Garden City, SG6 4ET, UK

³NeuroMET coordinator Milena Quaglia, LGC

Abstract. Consistent diagnosis in healthcare relies, in part, on quality assurance of categorical observations, such as responses to ability tests and patient surveys. Linking classifications on such nominal and ordinal scales to decision-making involves a combination of logit transformations and novel entropy-based estimates of measurement information throughout the measurement process. This paper presents how entropy can explain and predict entity attributes (such as task difficulty), instrument ability and resolution, and measurement system response. Cognitive ability studies in the EMPIR NeuroMET project are taken as an example, showing how better understanding of both entity and measurement system attributes leads to more fit-for-purpose and better targeted treatment.

1 Introduction

To diagnose, treat and rehabilitate consistently throughout the healthcare system requires reliable decision-making based on metrological quality assurance. However, important groups of observations in healthcare are considered to lie ‘off the scale’ of quantitative measurement. A fundamental reappraisal of metrology is needed if such ordinal or nominal properties are to be included in an extended quantity calculus on which the SI could be based.

‘Counts’ (non-negative integers) of, for example the number of pills in a bottle, have only recently been proposed as ‘quantities of dimension one in the SI’. Yet more qualitative are ‘counted fractions’ (bounded by zero and one), such as performance metrics for ability tests or patient surveys. Drawing superficial analogies in social measurements to traditional measurement instruments (e.g., a thermometer) in terms of simple response error, or merely calling a questionnaire an ‘instrument’, does not go far enough.

2 Categorical observations and decision-making. Entropy

Each measurement response is an elementary act of classification - identification or choice - into a particular category.

Misclassification probabilities have been proposed as accuracy measures for nominal examinations [1 – 3], but such performance metrics are ordinal and not directly amenable to regular statistics. The response of *Man as a Measurement Instrument* [4] at the heart of a measurement system is best handled instead with the Rasch, psychometric model in which

$$\theta - \delta = \log \left(\frac{P_{\text{success}}}{1 - P_{\text{success}}} \right) \quad (1)$$

gives test person ability, θ , and task difficulty, δ , by logistic regression to the score data in terms of the probabilities q (P_{success}) how well a measurement system, with *Man as a measurement instrument*, performs.

In this paper, we explore how the concept of entropy can be invoked to describe “dissipation of useful information” – to paraphrase Carnot [5] – at each of the three main stages in the measurement process [Figure 1] – from (A) an object’s entity, through (B) measurement with an instrument to (C) response as registered by an operator.

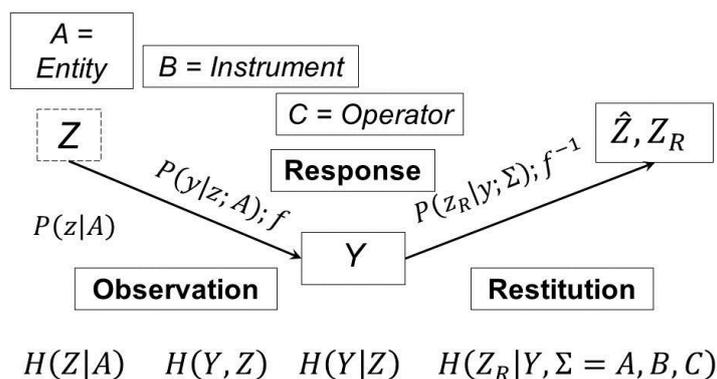


Figure 1. Probabilistic and entropy models of the measurement system and processes (inspired in part by the probabilistic model of Rossi [2]; Figure 5.5). Z – measurand; Y – response; R – restitution; P – probability; H – entropy

* Corresponding author: lesliependrill@gmail.com

An entropy-based approach allows not only a descriptive presentation (as in probability theory of measurement [2]) but also explanation and prediction. For instance, when describing the quality characteristic of the measured entity (measurand Z), a task will be easier if there is some degree of order, i.e., less entropy. Similarly, the response Y of a poorly performing measurement system can be explained in terms of both distortion and loss of information measured in terms of increases in entropy [6], i.e., disorder.

Typical measurement results derived from a Rasch analysis of (ordinal or nominal performance) test scores are illustrated in Figure 2, where probability mass functions (PMF) on a common and linear scale show the distribution of the ability among the individual test persons (upper, blue columns) and the distribution of task difficulty (lower, red columns) for each item of the classic Knox Cube Test (KCT) [7].

The distributions presented in Figure 2 reflect not only the task and person attributes intended to be measured, but also contain some aspects and limitations arising from imperfections in the measurement system employed to make the measurements which need to be compensated for.

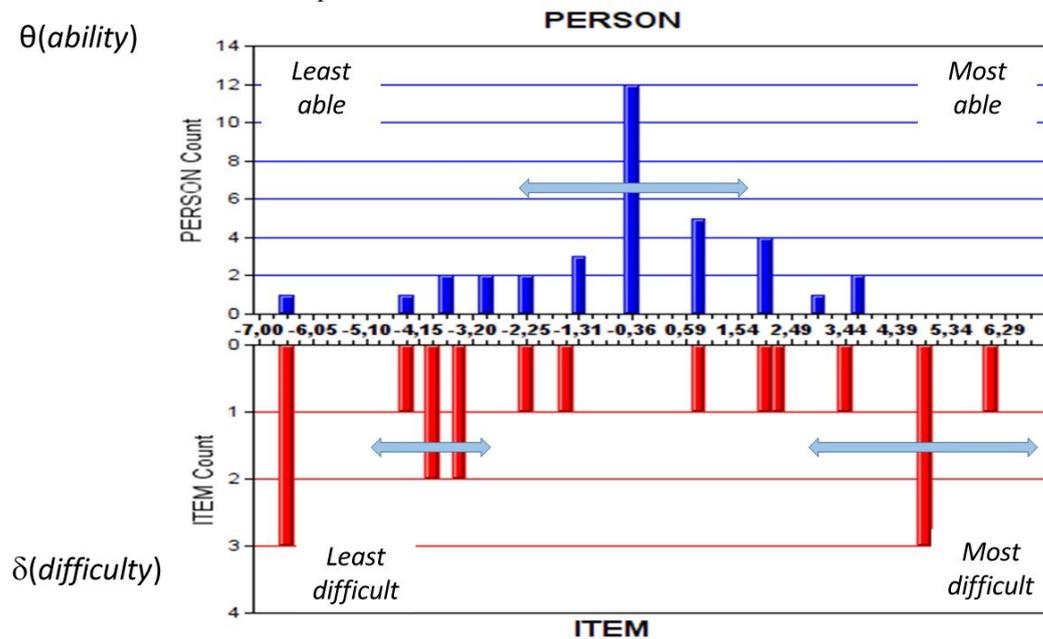


Figure 2. PMF distributions of test person ability, θ , (upper, blue columns) and task difficulty, δ , (lower, red columns) for the Knox Cube Test. ‘PERSON’ denotes test person. Uncertainty coverage factor $k = 2$ [Data from WINSTEPS® example ‘EXAM1.TXT’.]

3 Innovative measurement of neurodegenerative diseases

This new approach of logistic regression and entropy-based explanations is being applied in the on-going EMPIR HLT04 NeuroMET project *Innovative measurements for improved diagnosis and management of neurodegenerative diseases*. This is addressing a serious lack of metrologically sound assessment protocols and measurements for cognition, and a need for measurement comparability through the SI (International System of Units), traceability and uncertainty for regulatory approval of biomarkers for these diseases. Re-examination of traditional, widely used ‘legacy’ cognitive assessment protocols (e.g., MMSE, Corsi Block Test, Digital Span Test) using invariant measurement theory captures more accurately the patient’s cognitive ability and gives a better understanding of cognitive task difficulty [7].

3.1 Construct specification equations (CSE)

A CSE can be formulated for each construct (Y ; e.g. person ability or task difficulty) of interest as a linear combination of explanatory variables, X , (such as disease biomarker concentrations or test sequence length, respectively):

$$\hat{Y} = \sum_k \beta_k \cdot x_k \quad (2)$$

The CSE approach at the highest level of construct theory goes beyond a mere description to give a predictive tool for the design of for instance new cognitive tasks and abilities. Existing scales of cognitive difficulty can be complemented, and item equivalence demonstrated, thus permitting ‘tailor-making’ of novel cognitive assessment protocols. CSEs provide a means of formulating ‘recipes for certified reference materials’ in the social sciences [4], relating task difficulty to test construction [7] as well as provide a comprehensive understanding of person ability.

3.2 Explaining constructs with entropy

Ultimately the aim is to compensate as far as possible for the effects of imperfections in the measurement process, in order to obtain the most faithful measure of the quantities of interest: task difficulty and person ability, for instance.

The change in information step by step through our prototype measurement system [figure 1] can be expressed as a sum of entropy terms (exemplified below) of the well-known conditional entropy expression:

$$H(Q|P) = H(P, Q) - H(P) \quad (3)$$

where $Q = P(y|z)$ and $P = P(z)$. Eq. (3) simply states how the amount of information transmitted by a measurement system starts as an initial ‘deficit’ in entropy coming from prior knowledge, $H(P) = H(Z|A)$ of the measurand, to which are added losses and distortions $H(P, Q) = H(Y, Z)$ arising from imperfections which increase entropy during the measurement process.

Losses and distortions will mean that information is distributed over a range of categories, c , as depicted in figure 3. The total amount of measurement information on the categorical scales of signals at any one point between stimulus (Z) from the entity (A), through the instrument (B) and operator (C) to restitution of the measurement value (Z_R) from the response (Y), is in general the summed (increase in) entropy, which for a discrete PMF is $\Delta H(Q) = -\sum_c q_c \cdot \ln(q_c)$, where q_c is the occupancy of category c .

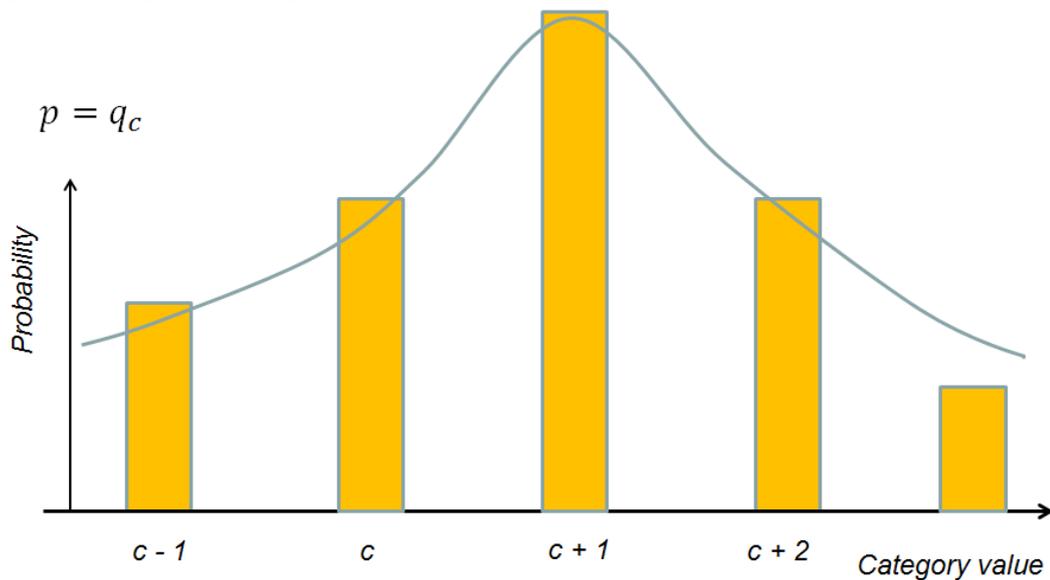


Figure 3. Entropy (amount of information): $\Delta H(Q) = -\sum_c q_c \cdot \ln(q_c)$ on a categorical scale

The situation where discrete multinomial expressions invoked for categorical responses go towards the familiar continuous scales of traditional uncertainty presentations can be expressed by:

$$\Delta H(Q) = -\int_{-\infty}^{\infty} p(Q) \cdot \ln(Q) \cdot dQ = \ln[\sqrt{2\pi} \cdot u(Q)] + \frac{1}{2} \quad (4)$$

that is, integrating the probabilistic (Shannon) formulation of the entropy across the width (‘uncertainty’) in the response. The last term on the RHS of eq. (4) applies to the particular case where the probability distribution function of the outcome Q is taken to be Gaussian (Normal), i.e. $p(Q) = N[\bar{Q}, u(Q)]$.

Inversion of Eq. (4) suggests an alternative expression of measurement uncertainty

$$u_q \sim e^{\Delta H(Q)} \quad (5)$$

more akin to the concepts of information theory than the classic standard uncertainties [JCGM GUM], indicating that the two approaches – (i) standard uncertainty, u , and (ii) decisions risks – can be unified. We have a certain preference to express uncertainty in terms of an increase ΔH in entropy instead of a standard deviation [8]: entropy is conceptually closer to ‘uncertainty’ in everyday language – ‘decision quandary’; is also substantially distribution-free; and is indeed is accessible to treatment not only with probability theory but also possibility and plausibility theories. In the following sections we explore analogous forms of eq. (5) throughout the measurement process.

3.2.1 A: Entity construct description and specification. Task difficulty

We start with the quality characteristic, δ , of the entity, A , of interest (product quality; task difficulty, etc) and how the concept of entropy can assist in describing, predicting and prioritising the entitic construct.

Consider a number, G , of categories (or cells) and a number, N_j , ($j=1, \dots, M$) of symbols of M different types. Brillouin’s [9] expression for the amount of information in messages of this kind can be used to develop specification equations for task difficulty for many sequence memory tests. Thus, apart from shorter sequences of tapped blocks in

the KCT (and the similar Corsi Block Test) being easier to memorise, words at the beginning and end of lists of the Auditory Verbal Learning Test (AVLT) are known to be easier to remember for most people than words in the middle, reflecting the effects of primacy and recency.

The task difficulty, $\delta = -[\ln(G!) - \sum_{j=1}^M \ln(N_j!)]$ of remembering a particular sequence of taps on a set of blocks in the KCT for cognitive memory function is one case study using the Brillouin expression [7]. A KCT sequence, such as: 1-4-2-3-4-1, can be expected to be easier to recall according to $-\sum_{j=1}^M \ln(N_j!)$ than a sequence of the same length but without repeats which has $-\ln(G!)$ information.

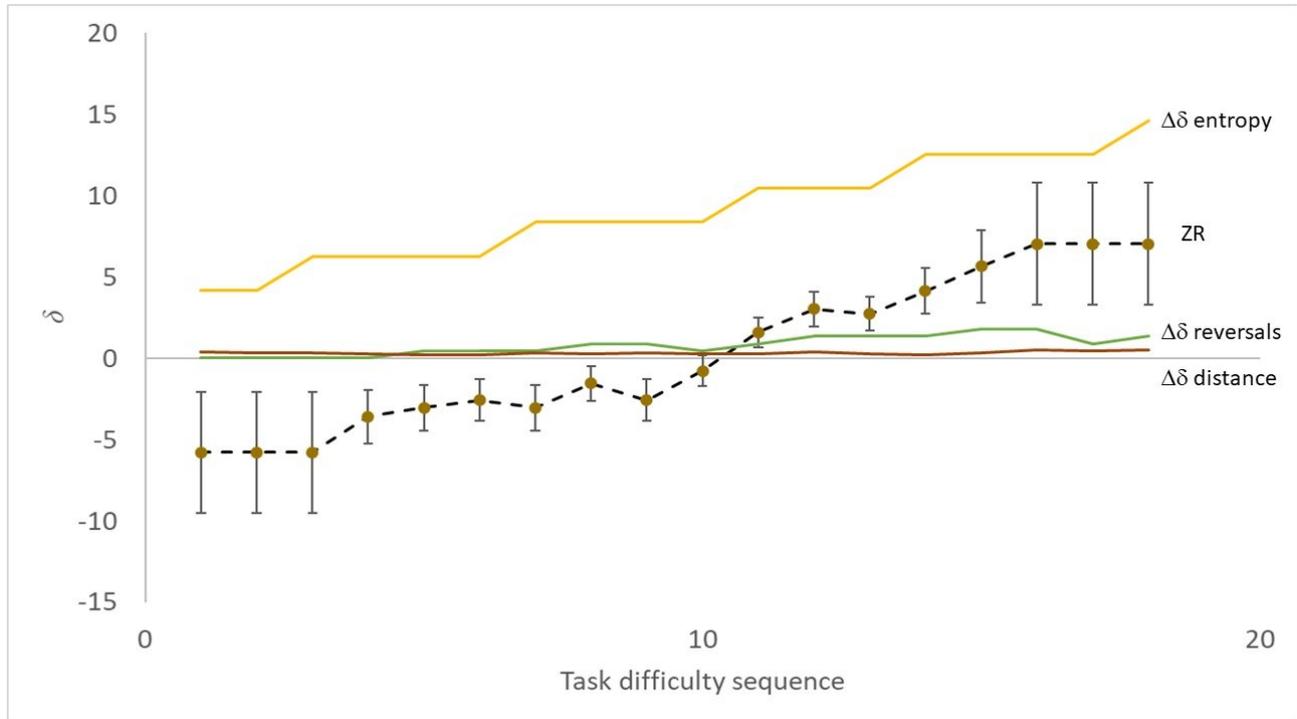


Figure 4. Predicted contributions, $\Delta\delta$, to task difficulty [zR (experiment)] from the three explanatory variables $X = \{\text{Entropy}, \text{Reversals}, \text{Average Distance}\}$ for the series of KCT sequences of increasing difficulty from a CSE based on entropy [eq. 6]

The KCT construct specification equation from principal component regression:

$$\delta_j = -9 (5) + 2 (1) \cdot \text{Entropy}_j + 0,8 (1,4) \cdot \text{Reversals}_j + 0,7 (2,9) \cdot \text{AveDistance}_j \quad (6)$$

was found to predict faithfully the experimentally observed task difficulties [7].

3.2.2 B: Instrument construct description and specification

Continuing the passage of information, in this section the role of the instrument at the heart of the measurement system is now considered and how the concept of entropy can assist in describing, predicting and prioritising the instrument construct [Figure 1].

There are many diverse examples of entropy-related changes in instrument performance. One can be found in recent neurological research, with a study of the entropy of interconnectedness amongst different regions of the brain [10]: the greater the order in brain processes, the smaller the entropy and the greater the (instrument) ability, for example to memorise. The same concept can also explain organisational efficiency in terms of entropy-based measures of synergy [11]. An explanation of the ability, θ , of an instrument (e.g., person) analogous to our explanation of task difficulty is in terms of an entropy term $\theta = H(P, Q) = -\ln(G!)$, where G is the number of ‘coherent connexions’ between different parts of the instrument.

Another term is the entropy $H(P, Q) \sim \ln(\rho) = \ln(\sqrt{3} \cdot 2 \cdot u)$ of a uniform distribution associated with the finite resolution, ρ , of an instrument similarly to eq. 5, where u is the standard measurement uncertainty.

The sum of these instrument-related entropies increases the initial stimulus entropy $H(P) = \delta$, so that the resulting response, P_{success} , is similar to what is described by the 2PL IRT expression [12]:

$$\ln \left[\frac{P_{\text{success}}}{1 - P_{\text{success}}} \right] = \theta - \delta + \ln(\rho) \quad (7)$$

where ρ is the discrimination,

One might suspect that there is some connexion between (person or instrument) ability, θ and instrument discrimination, ρ : the more able the instrument is, the greater the discrimination will be. Restitution of the measurement value from in this case a categorical ordinal or nominal response, is then proposed to be based on 2PL IRT, and takes the form of eq.7.

3.2.3 C: Response, error and entropy. Categorical observations

Finalising the passage of information, in this section the concept of entropy can also assist in describing, predicting and prioritising how the operator judges the response and performs final restitution of the measurand from the response of the measurement system [Figure 1].

Distortion of measurement information in general including ordinal and nominal data arising somewhere in the measurement system can be expressed as:

$$\text{Accuracy (decision-making)} = \text{response categorisation} - \text{input (true) categorisation} \quad (8)$$

Again entropy is invoked, this time in terms of the principle of maximum (Shannon) entropy, in which the change in entropy on transmission of measurement information from stimulus through response of the measurement system cannot decrease. Applying this principle and the Lagrange multiplier approach, maximizing the entropy function in response, leads to the categorical probability, q_c , of response: $q_c = \frac{e^{-\rho \cdot y_c}}{\sum_{c=1}^C e^{-\rho \cdot y_{ct}}}$ [4].

This maximum entropy derivation, together with eq. (7), in turn allows the formulation of novel construct specification equations for patient cognitive ability as a function of diverse biomarkers (e.g., in plasma, CSF and saliva together with MRI/MRS data) and brain structure. As a result, significant corrections are needed [Figure 5] to the classical test theory analyses still commonly done by leading neuroscientists [4]. Because of significant correlation amongst biomarkers, principal component regression is essential, as reported in the EMPIR NeuroMET project.

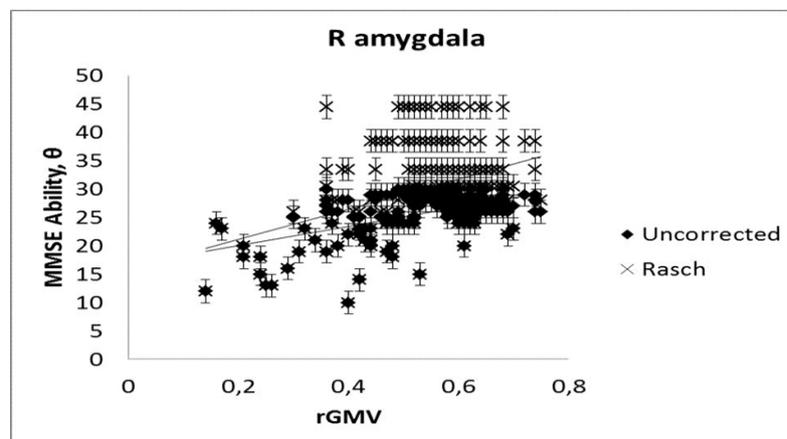


Figure 5. Correlation plots of cognitive ability (MMSE, dependent variable) versus regional grey matter volume (rGMV, independent variable) ♦ original data [13]; x Rasch corrected for ‘counted fraction’ distortion ([14] and [4]). Standard uncertainties are indicated for MMSE ability scores.

Conclusions

The deployment of the concept of entropy reported here leads to the realization of more fit-for-purpose, better targeted and better administered cognitive measurement systems. Traceable calibration is thereby enabled of both additional cognitive tasks as well as of the effects of intervention (or disease progression) on the cognitive ability of each individual patient.

Part of this work has been performed in the 15HLT04 NeuroMET project. This project has received funding from the EMPIR programme co-financed by the Participating States and from the European Union’s Horizon 2020 research and innovation programme.

References

1. E Bashkansky and T Gadrich 2010, “Some metrological aspects of ordinal measurements”, *AQUAL* **15**, 331 – 6, doi: 10.1007/s00769-009-0620-x
2. G B Rossi 2014, *Measurement and Probability – A probabilistic theory of measurement with applications*, Springer Series in Measurement Science and Technology, doi 10.1007/978-94-017-8825-0
3. G Nordin, *et al.* 2018, “Vocabulary on nominal property, examination, and related concepts for clinical laboratory sciences (IFCC-IUPAC Recommendations 2017)”, *Pure Appl. Chem.* **90**(5): 913–935 <https://doi.org/10.1515/pac-2011-0613>
4. L R Pendrill 2018, “Assuring measurement quality in person-centred healthcare”, *Measurement Science & Technology*, **29**, 034003 special issue Metrologie 2017, <https://doi.org/10.1088/1361-6501/aa9cd2>
5. L Carnot 1803, *Principes Fondamentaux de l'Équilibre et du Movement* [Fundamental Principles of Equilibrium and Movement], Paris
6. W. Weaver and C Shannon 1963, *The Mathematical Theory of Communication*. Univ. of Illinois Press. ISBN 0252725484
7. J Melin, *et al* 2019, ”Metrology of human-based perceptions: The role of entropy in construct specification equations to improve the validity of cognitive tests”, *Perspectives on Science (POS)*, Measurement at the Crossroads” special issue, submitted 190520
8. J V Zidek and C van Eeden 2003, ”Uncertainty, Entropy, Variance and the Effect of Partial Information”, *Lecture Notes-Monograph Series*, **42**, Mathematical Statistics and Applications: Festschrift for Constance van Eeden, pp 155 – 67
9. L Brillouin 1962, “Science and Information Theory”, *Physics Today* **15**, 2nd edition, Academic Press, DOI: 10.1063/1.3057866
10. Y. Yao, *et al* 2013, ”The Increase of the Functional Entropy of the Human Brain with Age”, *SCIENTIFIC REPORTS* | 3 : 2853 | doi:10.1038/srep02853, www.nature.com/scientificreports
11. Lei Chen, *et al.* 2015, “Collaborative Performance Research on Multi-level Hospital Management Based on Synergy Entropy-HoQ”, *Entropy* **17**, 2409 – 31, doi 10.3390/e1704240
12. I Partchev 2004, *A visual guide to item response theory*, Friedrich-Schiller-Universität Jena, <https://www.coursehero.com/file/28232270/Partchev-VisualIRTpdf/>
13. M Dinomais, *et a.l* 2016 “Anatomic correlation of the mini-mental state examination: a voxel-based morphometric study in older adults” *PLoS One* **11** e0162889
14. L F Hughes, *et al* 2003 “Using a Rasch scale to characterize the clinical features of patients with a clinical diagnosis of uncertain, probable, or possible Alzheimer disease at intake”, *J. Alzheimer's Dis.* **5** 367–73