ACTA IMEKO
ISSN: 2221-870X
December 2023, Volume 12, Number 4, 1 - 5

Entropy-based explanations of serial position and learning effects in ordinal responses to word list tests

Jeanette Melin[1], Petronella Kettunen[2, 3], Anders Wallin[2, 3], Leslie Pendrill[1]

[1] *RISE Research Institutes of Sweden, Division Safety and transport, Measurement Science and Technology Development, Gothenburg, Sweden*
[2] *Institute of Neuroscience and Physiology, Department of Psychiatry and Neurochemistry, University of Gothenburg, Gothenburg, Sweden*
[3] *Department of Psychiatry, Cognition and Old Age Psychiatry, Sahlgrenska University Hospital, Region Västra Götaland, Mölndal, Sweden*

ABSTRACT
Measuring a person's cognitive abilities, such as memory and learning, is central in many medical conditions to reliably diagnose, treat and monitor disease progression. Common tests typically include tasks of recalling sequences of blocks, digits or words. Recalling a word list is affected by so-called serial position effects (SPE), meaning that words at the beginning or end of the list are more likely to be recalled. In our earlier work, as part of including ordinal and nominal properties in metrology, compensation for ordinality in the raw test scores has been performed with psychometric Rasch measurement theory. Thereafter, SPE have been successfully explained with construct specification equations (CSE) dominated by information theoretical entropy as candidate reference measurement procedures. Here, we present how previous German results for explaining memory difficulty in the immediate recalling (IR, trial 1) task of the Rey's Auditory Verbal Learning Test (RAVLT) can be replicated with a Swedish cohort (the Gothenburg Mild Cognitive Impairment study, *n* = 251). This CSE replicability for RAVLT demonstrates comparability across the two cohorts in a kind of inter-laboratory study. Moreover, RAVLT includes repeated trials and learning through practice is expected. How memory task difficulty changes over the eight trials in RAVLT is studied: SPE are not so prominent for the delayed recalling sequences and there is an overall reduction in the task difficulty CSE intercept with trial number, interpreted as an effect of learning. To conclude, the methodology and evidence provided here can be clinically used not only to measure a person's memory ability but also his or her learning ability, as well as understanding the relationship between learning ability and other cognitive domains.

Section: RESEARCH PAPER

Keywords: Entropy; information; metrology; Rasch; cognition; memory; task difficulty; person ability; learning

Citation: Jeanette Melin, Petronella Kettunen, Anders Wallin, Leslie Pendrill, Entropy-based explanations of serial position and learning effects in ordinal responses to word list tests, Acta IMEKO, vol. 12, no. 4, article 29, December 2023, identifier: IMEKO-ACTA-12 (2023)-04-29

Section Editor: Eric Benoit, Université Savoie Mont Blanc, France

Received July 4, 2022; **In final form** December 12, 2023; **Published** December 2023

Funding: Part of this work has been done in the project 18HLT09 NeuroMET2, which 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.

Corresponding author: Jeanette Melin, e-mail: jeanette.melin@ri.se

1. INTRODUCTION

Worldwide, dementia is one of the most pressing public health issues in modern time [1]. At the same time, patients remain underdiagnosed or are diagnosed 'too late'. This can, in part, be explained by the limitations of routinely used methods for neuropsychological cognitive assessments [2], [3]. Key drawbacks are:

(i) the ordinal scale used to classify human responses to items assessing cognitive functioning; and
(ii) the need to separate the independent attributes – person ability and item difficulty – in human responses.

Ignoring these drawbacks leads to considerable risk of incorrect conclusions based on the cognitive assessments, subsequently impacting clinical decision-making.

While not traditionally a part of metrology, ordinal and nominal properties can indeed be considered both necessary, for example for dementia care [2], [3], and suitable [4]–[6] for metrological quality assurance, as is being increasingly recognised. Together with methods to both compensate for the ordinality and separate person and item attributes, so called Construct Specification Equations

(CSE) provide a more specific, causal, and rigorously mathematical conceptualization of item attributes (e.g., memory task difficulties) than any other kind of construct theory [4], advancing both the validity and metrological quality assurance of human based measurements [4]–[6].

The reliable and valid measurement of a person's memory ability is central to diagnosing, treating and monitoring disease progression in many medical conditions. Tests used to measure a person's memory ability typically include tasks such as recalling sequences of blocks, digits or words. An example of a word recall test is the Rey's Auditory Verbal Learning Test (RAVLT) [7], which is made up of a list of 15 unrelated words, in which the test person is asked in a trial to recall as many of these words as possible. This trial is repeated five times with the same words and order, followed by another list of 15 unrelated words (called a distractor list, trial 6), before returning to the original word list (trial 7), and finally once more, after a delay of 20 to 30 minutes (trial 8). As the same word list is repeated in RAVLT, learning through practice is expected.

Practice effects in cognitive tests are typically seen as a confounder in assessing the person's ability in an isolated cognitive domain [8], [9], e.g., memory, attention, or executive functions. On the other hand, it is well-known that practice is necessary for learning, and learning itself is a commonly evaluated cognitive domain of interest in neuropsychological assessments [10].

Another well-known effect in learning word lists is the serial position effect (SPE), which implies that words in the beginning (primacy region, Pr) or at the end (recency region, Rr) of the list are more likely to be recalled [11]. In the recent literature, SPE have been claimed to be potential diagnostic tools, e.g., [12].

In our previous work, we have successfully been able to mathematically explain, in terms of informational entropy, why words in Pr and Rr are easier to remember than items in the middle (middle region, Mr) in the RAVLT immediate recall (IR, i.e., first trial) [13]-[15]. In contrast to the RAVLT with 15 words with a fixed order on repeated trials 1 – 5, the word learning list (WLL) test included in the CERAD test battery has only 10 words and the word order changes with each of the three repeated trials. For this test, we have also shown that our entropy-based theory developed for RAVLT could also be successfully replicated for WLL CERAD trial 1 [16]. However, with only 10 words, SPE are expected to be less pronounced [11] and the effects of SPE are not as pronounced with repeated WLL trials as shown in this study of RAVLT [16]. That work was done in the European EMPIR NeuroMET2 project with the German version of the RAVLT. Elsewhere, we have also detailed presentations of how entropy may have a broad applicability when formulating CSE in general [4]–[6].

To the best of our knowledge, current analyses of measuring a person's learning ability are often limited by a widespread improper handling of the drawbacks (i) and (ii) mentioned above. For example, Moradi et al [4] define learning in RAVLT as the score of trial 5 minus the score of trial 1. Raw scores equal 1 for pass or 0 for fail, i.e., an ordinal scale [17], [18]. However, because of ordinality, such classifications have no numerical meaning and only serve to clearly indicate the ordered categories [8]. Linear separate measures for person and item attributes necessary for metrological legitimacy are only attainable by restitution from the raw scores through a logistic regression such as the dichotomous Rasch model [19].

By combining our previous work on RAVLT IR with the need for better methods to assess learning, we aim in this work to

(i) test if the previous German cohort results for explaining memory task difficulty in RAVLT IR can be replicated in a Swedish cohort; and
(ii) explore how memory task difficulty changes over repeated trials of RAVLT.

2. METHODS

2.1. Participants and data collection

The Gothenburg (GBG) mild cognitive impairment (MCI) study is an attempt to conduct longitudinal, in-depth phenotyping of patients with different forms and degrees of cognitive impairment using neuropsychological, neuroimaging, and neurochemical tools [20]. The GBG MCI study was started in 1999 and has bi-annual visits.

For this study, we included a subsample with data from 251 individual assessments. The cohort is dominated by patients with MCI (n = 183) and healthy controls (n = 57), although some have had a disease progression from the second visit (Alzheimer's disease n = 3; vascular dementia n = 3; mixed type dementia [combined Alzheimer's disease and vascular dementia]) n = 3; dementia non-ultra descriptum n = 2). Mean age was 65 years (Std dev. 8) and 124 were men and 127 women.

2.2. Data analysis

A first step in handling the drawbacks (i) and (ii) mentioned above properly was to make a measurand restitution. Specifically, linear separate measures for person and item attributes suitable for metrological quality assurance were restituted through a logistic regression of the dichotomous Rasch model [19] to the ordinal response raw score using the software WINSTEPS ® 5.2.0. The Rasch model yields separate and linear measures for each memory task difficulty, δ, and for each individual person memory ability, θ, and compensates for ordinality:

$$P_{\text{success}} = \frac{\mathrm{e}^{\theta-\delta}}{1+\mathrm{e}^{\theta-\delta}}\,. \tag{1}$$

The focus of this study is primarily on measures of memory task difficulties, δ.

We used a so-called racked data entry, which means that items from the different trials are treated as individual items. This approach was chosen as it can be used to assess how item task difficulty changes over time and to compare numerically within the same frame of reference [21].

Secondly, CSE [4], [5] for memory task difficulty for each trial were formulated. Recently, we have provided an extensive description of the metrological significance of this and how this applies for word-learning list tests [15] (in particular, see Appendix B). In short, our approach is based on information theoretical entropy – a more ordered task with less entropy is expected to be easier [4]. The amount of information, I, in a message containing G symbols with N repeats of M_N different types was formulated in the work of Brillouin [22], based on the well-known Shannon [23] expression of 'surprisal':

$$I = M \cdot \left[\ln(G!) - \sum_{j=1}^{M_N} \ln\left(N_j!\right)\right]. \tag{2}$$

In the present case, the normalisation constant, $M = 1/\ln(L)$ $=1/\ln(15) = 0.369$ for a word list of length, $L = 15$. In our recent work [15], the following definitions for explanatory variables for SPE were presented:

$$\delta_{\mathrm{Mr},0} = 2 \cdot M \cdot \ln\left(G_j!\right) \; ; \; G = L/2 \,, \tag{3}$$

$$\delta_{\mathrm{Pr},j} = -M \cdot \ln\left(G_j\,!\right); \; G \; = \; item\,position, j \tag{4}$$

$$\delta_{\mathrm{Rr},j} = -M \cdot \ln\left(G_j\,!\right); \; G \; = \; L \; - 1 - \; item\,position, j \tag{5}$$

$$\delta_{\mathrm{freq},j} = -M \cdot \ln f_j \,. \tag{6}$$

Based on Equation (2), the average entropy at the middle of a list of G words, without SPE or repeats, is given by Equation (3). But due to the well-known effects of primacy and recency when all other factors being equal, the initial and final symbols in the word list should be somewhat easier to recall than the symbols in the middle of the sequence Equations (4) and (5). Word familiarity might be another effect that may explain the difficulties of recalling words from a list, Equation (6).

A final step in the formulation of a CSE for task difficulty is to account for cases where the principal components of variation are not the observed explanatory variables. This was done in a principal component regression (PCR), including three steps:

(i) A PCA amongst the set of explanatory variables, $\boldsymbol{X_k}$

(ii) A linear regression of the empirical task difficulty values δ_j against $\boldsymbol{X' = X \cdot P}$ in terms of the principal components, $\boldsymbol{P}$, and

(iii) A conversion back from principal components to the explanatory variables, $\boldsymbol{X_k}$

3. RESULTS

3.1. RAVLT IR replication of previous results

When comparing the empirical task difficulty values for RAVLT IR in this study (blue dots in Figure 1) with our previous work [13]-[15], similar CSE were observed. Likewise, when comparing the CSE based on the two cohorts, differences were found to be negligible (numbers in brackets indicate measurement uncertainties; coverage factor k = 2):

NeuroMET RAVLT

$$zR_{\mathrm{Trial}\;1,j} = 5(3) + 0.7(5) \cdot \delta_{\mathrm{Pr}\,j} + 0.8(5) \cdot \delta_{\mathrm{Rr}\,j} + 0.2(2) \cdot \delta_{\mathrm{freq},j} \tag{7}$$

GBG MCI RAVLT

$$zR_{\mathrm{Trial}\;1,j} = 5(3) + 0.6(4) \cdot \delta_{\mathrm{Pr}\,j} + 0.6(4) \cdot \delta_{\mathrm{Rr}\,j} - 0.2(2) \cdot \delta_{\mathrm{freq},j} \,. \tag{8}$$

3.2. RAVLT learning effects

With repeated trials, item memory task difficulty was found to decrease. This is shown in Figure 1; the blue dots for trial 1 illustrate that those items are more difficult to recall compared with the same words in the subsequent trials.

When comparing the CSE from trial 1, Equation (8), with the CSE from trials 2-5, Equations (9)-(12), all explanatory variable coefficients stayed stable within the measurement uncertainties: GBG MCI RAVLT

$$zR_{\mathrm{Trial}\;2,j} = 5(2) + 0.6(3) \cdot \delta_{\mathrm{Pr}\,j} + 0.7(5) \cdot \delta_{\mathrm{Rr}\,j} + 0.1(2) \cdot \delta_{\mathrm{freq},j} \tag{9}$$

$$zR_{\mathrm{Trial}\;3,j} = 4(2) + 0.6(3) \cdot \delta_{\mathrm{Pr}\,j} + 0.7(4) \cdot \delta_{\mathrm{Rr}\,j} + 0.3(1) \cdot \delta_{\mathrm{freq},j} \tag{10}$$

$$zR_{\mathrm{Trial}\;4,j} = 4(2) + 0.5(3) \cdot \delta_{\mathrm{Pr}\,j} + 0.6(4) \cdot \delta_{\mathrm{Rr}\,j} + 0.2(1) \cdot \delta_{\mathrm{freq},j} \tag{11}$$

$$zR_{\mathrm{Trial}\;5,j} = 3(2) + 0.5(3) \cdot \delta_{\mathrm{Pr}\,j} + 0.7(5) \cdot \delta_{\mathrm{Rr}\,j} + 0.3(2) \cdot \delta_{\mathrm{freq},j} \,. \tag{12}$$

An overall decrease in task difficulty is evident in the successive lowering of the repeated trial curves shown in Figure 1, which can be interpreted as a learning effect. Although uncertainties are large, this can be expressed by a lowering of the intercept (first term on the right-hand side (RHS) of each CSE for task difficulty) which was found to decrease significantly on each repeated trial and approximately linearly with trial number, as shown in Figure 2.

The reduction in task difficulty intercept, interpreted as learning, can be simply explained with the formula:

$$\delta_{\mathrm{intercept}}(trial) = \frac{\delta_{\mathrm{Mr},0}}{\sqrt{trial}} \,, \tag{13}$$

where $\delta_{\mathrm{Mr},0}$ is given by Equation (3) and *trial* is the number of each trial in order, by analogy to a reduction in uncertainties with increasing numbers of degrees of freedom as trials are repeated.

Memory task difficulty from trials 1, 5, 7 and 8 are compared in Figure 3. While trial 1 and 5 both have a parabolic relation between difficulty and item order, reflecting SPE (as in Figure 1), this is missing from trials 7 and 8. Thus, SPE are not as prominent for the later, delayed recalling sequences; words from *Pr* are easier to recall than words from *Rr*.

The differences in task difficulty between these trials are also evident when comparing the CSE. In Equations (14) and (15), the contributions from SPE are more or less negligible. In turn,

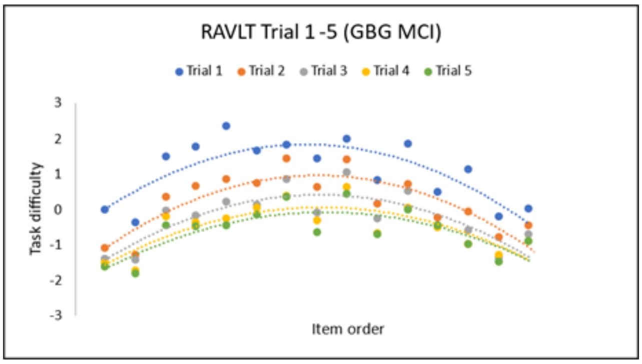

Figure 1. Item memory task difficulty on the y-axis and item position on the x-axis showing SPE (parabolic curves) and reductions in task difficulty of repeated trials.

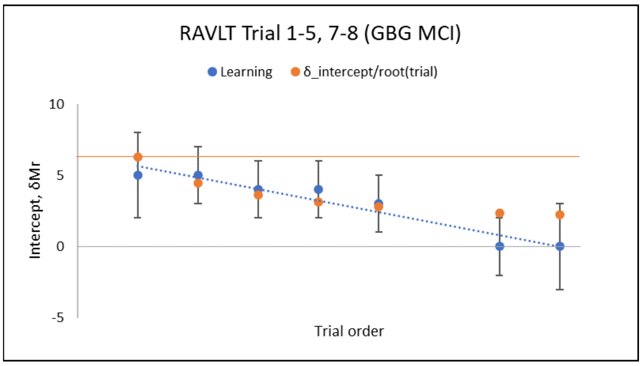

Figure 2. Item memory task difficulty intercept on the y-axis (Learning: experiment; δ_intercept: Equation (13) and trial on the x-axis showing the effects of learning.

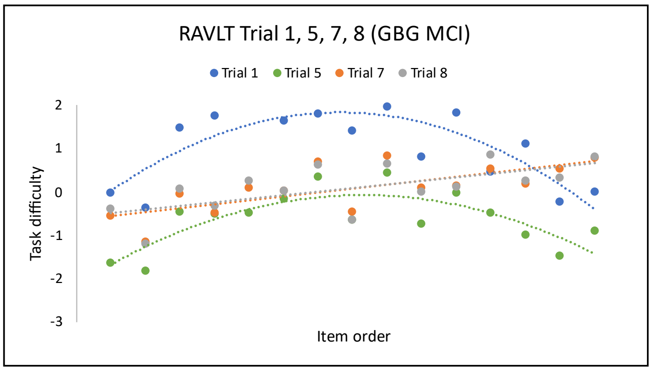

Figure 3. Item memory task difficulty on the y-axis and item position on the x-axis. SPE (parabolic curves) for the earlier trials (1 & 5) disappear for the later trials (7 & 8).

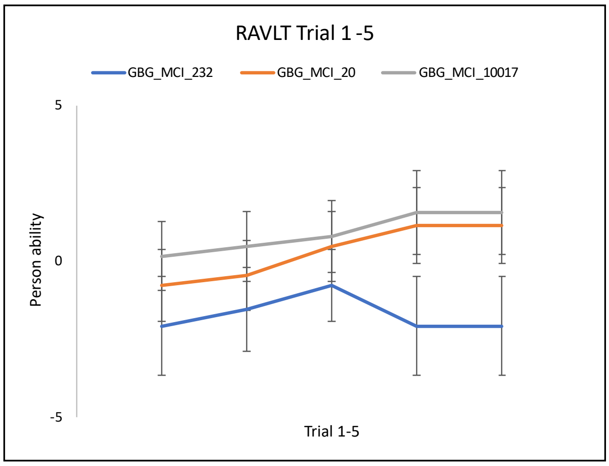

Figure 4. Person memory abilities on the y-axis and trial position on the x-axis. Error bar corresponds to measurement uncertainties k = 2.

the predictive power was found to decrease over the trials. When correlating the empirical task difficulty values with zR, the Pearson correlation coefficient was 0.80 for trial 1 but had reduced to 0.67 by trial 8.

GBG MCI RAVLT

$$zR_{\text{Trial } 7,j} = -0.2(2.0) + 0.0(3) \cdot \delta_{\text{Pr}_j} + 0.2(4) \cdot \delta_{\text{Rr}_j} + 0.2(1) \cdot \delta_{\text{freq},j} \quad (14)$$

$$zR_{\text{Trial } 8,j} = -0.2(2.6) + 0.0(3) \cdot \delta_{\text{Pr}_j} + 0.2(5) \cdot \delta_{\text{Rr}_j} + 0.2(2) \cdot \delta_{\text{freq},j} \, . \quad (15)$$

4. DISCUSSION

Inter-laboratory studies are well-established as a means of evaluating measurement accuracy and ensuring metrological traceability, for instance in chemical and materials metrology. In the human and social sciences, including neuropsychological assessments, such routines are less developed. This hinders ensuring true metrological traceability for individual persons as well as between cohorts and over time. In what can be seen as a kind of inter-laboratory study, the reproducibility of CSE between the German and Swedish versions of RAVLT presented here demonstrates comparability between the cohorts in Berlin and Gothenburg.

Our findings when explaining the evolution of SPE over the trials give arguably a clearer picture of differences between immediate and delayed recall than previously reported with classical test theory, e.g., [24].

In turn, this may have implications when claiming SPE as potential diagnostic tools. In this study, as well as in the WLL CERAD-study [16], the reduction in task difficulty intercept were both found to decrease in inverse proportion to the root of the number of trials performed. However, in the WLL CERAD study, we also found a difference in how fast the intercept decreases between groups of less or more cognitive able cohorts. This might indicate a faster learning for the more cognitive able cohort members. In a previous study by Cordier et al [25], in which entropy was used as a 'global variable' for learning, it was found that experts exhibited a faster decline in entropy and came earlier to a plateau. In the present study, with less diverse cognitive status of participants, we have not, as yet, been able to study such learning trajectories for different skilled groups, thus further studies are warranted on subjects within the full spectrum of AD also when using RAVLT.

Moreover, SPE may introduce a multidimensionality effect, which could seriously challenge the basic assumptions behind the metrological Rasch model, particularly its specific objectivity, as would arise where significant portions of the cohorts appear to experience SPE differently [14], [15]. However, in the present study, we have not been able to observe such a breakdown. This may be due to the less diverse cognitive status of participants, compared with our earlier work where cohort members ranged from health controls to patients suffering from dementia due to suspected Alzheimer's Disease [13]-[15].

Shifting focus from assessments of item task difficulty to how individual person ability varies across the trials may, however, can tell us something about different learning trajectories for further studies. For example, Figure 4 show person abilities for three cohort members for RAVLT trial 1 to 5. Two persons, starting at trial 1 from slightly different abilities, both increase in ability to reach a comparable plateau at the fourth trial while a third person reaches a lower plateau at the fourth trial.

When items in trial 1 in RAVLT are scaled together with other items from short-term memory tests such as Corsi Block Test (CBT) and Digit Span Test (DST), RAVLT task difficulties are relatively uncertain. One can get a better precision compared to CBT and DST if used for the persons with abilities around 0.00 logits [26]. However, if RAVLT is being used to measure people with lower or higher abilities, the precision decreases. When only using the first trial with only 15 words might not be sensitive enough from a clinical perspective. Thus, further work is needed to handle both SPE and learning effects, – potentially causing multidimensionality and a breakdown in the Rasch model. More consistent ways of defining and analyzing memory task difficulty to maintain the unique metrological properties of the Rasch model are needed [15].

There are some further general methodological limitations to CSEs and entropy-based models. Only when all potentially important explanatory variables are included can a CSE give a true picture [6]. Thus, when formulating CSEs effects of sample size, collinearity, a measurement disturbance, and multidimensionality on the estimation of component difficulties must be considered [27]. However, despite such potential limitations, the benefits are more important, specifically, the lower measurement uncertainties in the multivariate models compared with univariate fits are obtained which is of great significance in the present field of memory measurements, and beyond [6].

5. CONCLUSION

To conclude, we have presented both a methodology [25] to overcome current practice shortcomings in analyzing word learning list tests, i.e., using the Rasch model [17] and CSE [4], [5], as well as additional evidence for how repeated trials can lead to learning effects. This can be clinically used not only to measure a person's memory but also his or her learning ability. Further work is needed to find more consistent ways of defining and analyzing memory task difficulty, including SPE and learning effects, to maintain the unique metrological properties of the Rasch model and to improve the estimates and understanding of person memory abilities [26]. In turn this should enable a better understanding of the relationship between learning ability and other cognitive domains as well as more reliable diagnosing, treating, and monitoring of disease progression.

REFERENCES

[1] European Brain Council, Driving policy to optimise care for people with Alzheimer's Disease in Europe today and tomorrow, 2018. Online [Accessed 18 December 2023] https://www.braincouncil.eu/wp-content/uploads/2018/11/Driving-policy-to-optimise-care-WEB.pdf

[2] J. Hobart, S. Cano, H. Posner, O. Selnes, Y. Stern, R. Thomas, J. Zajicek, Alzheimer's Disease Neuroimaging Initiative. Putting the Alzheimer's cognitive test to the test II: Rasch measurement theory, Alzheimers Dement. (2013). 9, S10–20. DOI: 10.1016/j.jalz.2012.08.006

[3] J. Hobart, S. Cano, H. Posner, O. Selens, Y. Stern, R. Thomas, J. Zajicek, Alzheimer's Disease Neuroimaging Initiative. Putting the Alzheimer's cognitive test to the test I: Traditional psychometric methods, Alzheimers Dement. (2013) 9, S4–9. DOI: 10.1016/j.jalz.2012.08.005

[4] J. Melin, S. Cano, L. Pendrill, The role of entropy in construct specification equations (CSE) to improve the validity of memory tests, Entropy 2021, 23 (2), 212. DOI: 10.3390/e23020212

[5] L. Pendrill, Quality assured measurement: Unification across social and physical sciences. Springer International Publishing. (2019). DOI: 10.1007/978-3-030-28695-8

[6] J. Melin, L. Pendrill, The role of construct specification equations and entropy in the measurement of memory, In: Fisher Jr. William P., Cano S.J., eds. Person-centered outcome metrology: Principles and applications for high stakes decision making. Cham. Springer International Publishing. (2023) pp. 269–309. DOI: 10.1007/978-3-031-07465-3_10

[7] A. Rey, L'examen clinique en psychologie. [The clinical examination in psychology.] Oxford, England: Presses Universitaries De France (1958). [In French]

[8] T. E. Goldberg, P. D. Harvey, K. A. Wesnes, P. J. Snyder, L. S. Schneider, Practice effects due to serial cognitive assessment: Implications for preclinical Alzheimer's disease randomized controlled trials, Alzheimers Dement. Diagn. Assess. Dis. Monit. (2015) 1, pp. 103–11. DOI: 10.1016/j.dadm.2014.11.003

[9] K. A. Hawkins D. Dean, G. D. Pearlson, Alternative Forms of the Rey Auditory Verbal Learning Test: A Review, Behav. Neurol. (2004)15, pp. 99–107. DOI: 10.1155/2004/940191

[10] E. Moradi, I. Hallikainen, T. Hänninen, J Tohka, Rey's Auditory Verbal Learning Test scores can be predicted from whole brain MRI in Alzheimer's disease, NeuroImage Clin. (2017) 13, pp. 415–27. DOI: 10.1016/j.nicl.2016.12.011

[11] B. B. Murdock Jr., The serial position effect of free recall, J. Exp. Psychol. (1962) 64, pp. 482–8. DOI: 10.1037/h0045106

[12] D. S. Weitzner, M. Calamia, Serial position effects on list learning tasks in mild cognitive impairment and Alzheimer's disease, Neuropsychology (2020) 34, pp. 467–78. DOI: 10.1037/neu0000620

[13] J. Melin, A. Regnault, S. Cano, L. Pendrill, Neuropsychological assessments: Word learning tests and diagnostic potential of serial position effects, The International Metrology Congress, Lyon, France, 7-9 September 2021.

[14] L. Pendrill, J. Melin, S. Cano, Entropy-based explanations of multidimensionality in ordinal responses, MSMM, Joint Workshop of ENBIS and MATHMET, Virtual.

[15] J. Melin, S. Cano, A. Flöel, L. Göschel, L. Pendrill, The role of entropy in construct specification equations (CSE) to improve the validity of memory tests: Extension to word lists, Entropy, (2022) 24 (7). DOI: 10.3390/e24070934

[16] J. Melin, L. Pendrill, A Novel Metrological Approach to a More Consistent Way of Defining and Analyzing Memory Task Difficulty in Word Learning List Tests with Repeated Trials, RaPID4, Marseille, France, 2022.

[17] V. Turetsky, E. Bashkansky. Ordinal response variation of the polytomous Rasch model, METRON 2022, 80 (3), pp. 305–33. Online [Accessed 18 December 2023] https://link.springer.com/content/pdf/10.1007/s40300-022-00229-w.pdf

[18] S. S. Stevens, On the Theory of Scales of Measurement, Science, New Series, 1946, 103(2684), pp. 677–680. DOI: 10.1126/science.103.2684.677

[19] G. Rasch, Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests, Oxford, England: Nielsen & Lydiche (1960).

[20] A. Wallin, A. Nordlund, M. Jonsson, K. Lind, Å. Edman, M. Göthlin, J. Stålhammar (+ 12 more authors), The Gothenburg MCI study: Design and distribution of Alzheimer's disease and subcortical vascular disease diagnoses from baseline to 6-year follow-up, J. Cereb. Blood Flow Metab. (2016) 36, pp. 114–31. DOI: 10.1038/jcbfm.2015.147

[21] B. Wright, Rack and Stack: Time 1 vs. Time 2 or Pre-Test vs. Post-Test, Rasch Meas. Trans. (2003) 17 (1), pp. 905–906. Online [Accessed 18 December 2023] https://www.rasch.org/rmt/rmt171a.htm

[22] L. Brillouin, Science and information theory, New York. (1962)

[23] C. Shannon, A mathematical theory of communication, Bell System Technical Journal, (1948) 27, pp. 379-423. DOI: 10.1002/j.1538-7305.1948.tb01338.x

[24] C. Turchetta, M. De Simone, R. Perri, L. Fadda, G. Caruso, M. De Tollis, C. Caltagirone, G. Carlesimo, Forgetting rates on the recency portion of a word list predict conversion from Mild Cognitive Impairment to Alzheimer's Disease, J. Alzheimers Dis. (2020) 73, p. 1295–304. DOI: 10.3233/JAD-190509

[25] P. Cordier, M. M. France, J. Pailhous, P. Bolon, Entropy as a global variable of the learning process, Hum. Mov. Sci. (1994) 13 (6), pp. 745–763. DOI: 10.1016/0167-9457(94)90016-7

[26] J. Melin, S. Cano, A. Flöel, L. Göschel, L. Pendrill, Construct specification equations: 'Recipes' for certified reference materials in cognitive measurement, Meas. Sens. (2021) 18, 100290. DOI: 10.1016/j.measen.2021.100290

[27] K. E. Green, R. M. Smith, A Comparison of Two Methods of Decomposing Item Difficulties, Journal of Educational Statistics, Vol. 12, No. 4 (Winter, 1987), pp. 369-381 (13 pages). DOI: 10.2307/1165055