Abstract

Anagrams are frequently used by experimental psychologists interested in how the mental lexicon is organised. Until very recently, research has overlooked the importance of syllable structure in solving anagrams and assumed that solution difficulty was mainly due to frequency factors (e.g. bigram statistics). The present study uses Rasch analysis to demonstrate that the number of syllables is a very important factor influencing anagram solution difficulty for both good and poor problem solvers, with multi-syllable words being harder to solve. Furthermore, it suggests that syllable frequency may have an impact on solution times for multi-syllable words with more frequent syllables being more difficult to solve. The study illustrates the advantages of Rasch analysis for reliable and unidemensional measurement of item difficulty.

Keywords: Cognition, Problem-solving, Rasch analysis, Individual differences

The role of syllables in anagram solution: A Rasch analysis

There is a long history of anagrams being used in experimental psychology as tools to investigate cognitive processes. However, there is still uncertainty as to the factors which influence anagram difficulty. Until recently, the tacit assumption appears to be that anagram difficulty is largely a function of word frequency. As reliable and objective measurement is crucial, this paper sets out to apply Rasch scaling to a sample of five-letter anagrams to examine whether a single unidimensional scale based on syllable number can be usefully applied. Manipulating anagram difficulty reliably is important where they are used to induce anxiety (e.g. Endler, Speer, Johnson & Flett, 2001) or cognitive load (e.g. Beversdorf et al. 2007; Foley & Foley, 2007).

Experimental studies that have explored individual differences in anagram solving have provided insights into aspects of human reasoning and problem-solving (e.g. Novick & Sherman, 2003). Novick and Sherman (2008) reported two experiments that showed the importance of the number of syllables in a word on its solution time, when it was presented as a five-letter anagram. Overall they found that two-syllable anagrams took longer to solve than one-syllable anagrams. Further they found that this effect was particularly marked for good anagram solvers. This result is rather surprising as in over fifty years of research on anagram solving, no other study has found or even suggested the possibility of a syllable effect on anagram solution. Word frequency, age-of-acquistion of the word, its meaningfulness, concreteness and imagery among many other attributes have been suggested as factors that influence solution (for example, Gilhooly & Johnson, 1978), but not number of syllables.

As with all research, it is important to demonstrate that Novick and Sherman's finding is not unique to the stimuli, the participants chosen and the method of investigation used. In this case it is particularly salient as Novick and Sherman (2008) compared two groups of good and poor solvers who were selected on the basis of their ability to solve "difficult" anagrams. The majority of these screening anagrams (70 per cent) were two-syllable anagrams, so it is possible that their results are related to selecting participants who are particularly sensitive to syllable effects on anagram solution. Furthermore, as has been commented on (Coleman, 1964) and demonstrated before (Clark, 1973), that generalizing the results of language experiments is particularly problematic as a significant result tells us only that the result is likely to generalize to a new set of participants and not necessarily to a new set of stimuli. It is also important to demonstrate that the syllable effect generalises to other indices of anagram difficulty than that used by Novick and Sherman.

In the present study we attempted to confirm the syllable effect found by Novick and Sherman (2008), using a different method of calculating solution difficulty, on a different set of anagrams, with a group of participant not selected by ability. We also included additional explanatory variables (e.g. *) and followed Gilhooly and Johnson's (1978) regression analysis approach to anagram solution. In their study of five-letter anagrams, 45 participants were given eighty anagrams to solve and the number of participants who solved an anagram was used as a dependent variable (i.e. index) of anagram difficulty. Gilhooly and Johnson (1978) then investigated the relative importance of twelve independent variables on solution score.

We used a regression method similar to that used by Gilhooly and Johnson (1978), but we also included a measure of competence of anagram solution and some new independent variables. Novick and Sherman (2008) measured competence with a 20 anagram screening test of difficult anagrams, 14 of which were two-syllable anagrams. In place of a pretest, Rasch analysis was used in this study to establish a participant's ability to solve anagrams and also to establish the relative difficulty of each anagram. Rasch analysis allows both person and item (anagram) parameters to be considered separately which allows us to consider relative competence in anagram solution without artificially creating a good and poor group of solvers. Rasch analysis also permits the investigation of how well a dependent variable, in this case anagram difficulty, meets the criterion of being both unidimensional and reliable and creates interval level data.

The basics of the analysis are as follows. To take the anagram information first; it is easy to work out the difficulty of each anagram by using the percentage of the sample of participants who get the answer correct. This can be transformed into the probability of getting the anagram correct or the odds of getting an item correct. We can also calculate the ability of each participant by taking the percentage of anagrams that they get correct and can then turn this into a probability of that person solving an anagram correctly. Rasch's theory suggests that the probability of getting an individual item (anagram) correct is caused by the difference in a person's ability and the item (anagram) difficulty. To put it simply if a person's ability is higher than a particular anagram's difficulty then the participant is more likely to get this correct than if it is lower than the anagram's difficulty. Using this information we can compare the data collected with what we would expect based on calculations of anagram difficulty and person ability. The closer the observed results are to the predicted results the better fit the data are to the Rasch model.

We included all of the variables examined by Gilhooly and Johnson (1978) in their analysis with the addition of two new variables related to syllables. The first is number of syllables which is similar to that used by Novick and Sherman (2008). We also included syllable frequency, as Stenneken, Conrad and Jacobs (2007) and Macizo and van Petten (2007) have recently shown a syllable frequency effect in lexical decision tasks.

Method

Participants

In total, 128 undergraduate students from the University of Durham participated in this study over two sessions. The first session involved 63 first-year Psychology undergraduates, the other 65 second-year Psychology undergraduates.

Design & Materials

The study was a within-participants design. All participants were presented with 80 five-letter anagrams (see appendix I) which they were required to solve. The words from which the single-solution anagrams were constructed were selected at random from the list of 205 nouns provided by Gilhooly and Hay (1977). Two- or three-move anagrams were constructed at random for each of the words. An example of a two-move anagram is HWTCA: WATCH. In total there were 51 two-move anagrams and 29 three-move anagrams. None of the words were plurals, proper names or had repeated letters.

Gilhooly and Johnson (1978) included the following twelve variables in their analysis; imagery, similarity, pronounceability, familiarity, concreteness, age-of-acquisition, meaningfulness, log of bigram rank, number of vowels, starting letter, GTZERO, and the log of word frequency. Most of these measures are self-explanatory, however, log of bigram ranks and GTZERO, both of which come from the bigram frequency matrix probably need some explanation. The bigram frequency matrix is constructed by drawing a table with 20 rows representing the 20 possible bigrams (two letter sequences) and four columns representing the four bigram positions in a five-letter word. The bigram rank is the number of entries in the table which have higher frequencies than the four correct entries (i.e. real bigram positions). GTZERO is also calculated from the bigram frequency matrix and is the total number of bigrams with a frequency of greater than zero in the bigram frequency matrix. For example, for the anagram IGTHL (Light) HG, HT, HL, GT, TG, TL, LH, LG, LT would all have a frequency of 0 in the first position. The more non-zero entries there are, the greater the possible competing solutions which make the anagram harder to solve (Mendelshon, 1976). It is conceptually similar to Ronning's (1965) "rule out factor" in which certain bigram possibilities are ruled out from consideration as they do not exist in the English language in certain positions.

We used the same measures for our anagrams from Gilhooly and Hay (1977), Gilhooly and Johnson, 1978), with the exception of pronounceability of the anagram which depends on the order of letters. We used the same method as Gilhooly and Johnson (1978) to measure pronounceability, by asking 16 adults to rate the pronounceability of a list of nonwords (i.e. the anagrams) using a 7-point scale (1 = "unpronounceable", and 7 = "very easy to pronounce"). The effective reliability of the pronounceability ratings for this study was $R = .97$.

As well as using the *Kucera-Francis* (1967) word frequency score which was used by Gilhooly and Johnson (1977) we also obtained objective frequency ratings from *HALfreq* (Balota, Cortese, Sergent-Marshall, Spieler & Yap 2004). In addition, we included frequency measures from the Thorndike and Lorge (1944) word count as this has been used in a number of other older anagram studies. We also obtained subjective frequency (Balota, Pilotti & Cortese, 2001) ratings from a sample of 26 second-year undergraduates using a 7-point scale, ranging from "never encountered", to seen "several times a day".

Number of syllables was determined by using the English lexicon project (Balota et al., 2002). Positional syllable frequencies were derived from the English orthographic wordform database of *CELEX*, which includes frequencies from a combined written and spoken corpus of 17.9 million words (Baayen, Piepenbrock & Gulikers, 1995). The orthographic syllabification was found for each wordform in the database, excluding proper nouns, abbreviations, and multi-word phrases. For each included wordform, the frequency of the wordform was summed with a two-dimensional table indexed by both the text of the syllable and its ordinal position in the wordform. The frequency of each syllable in the stimulus words was then found by looking up the word's syllabification and noting the relevant table entries. Previous research (Macizo & van Petten, 2007) suggests that the first syllable frequency will be the most important so we included the log of this frequency.

Procedure

The anagrams were presented across the two group sessions using the same procedure. The stimuli were presented via PowerPoint projection to the front of the class using the format of yellow letters (Arial Black 66 point-font) on blue background. Each anagram was shown for 15 seconds, with an inter-trial interval of 5 seconds. The participants were provided with a response sheet with numbered spaces in which to write their answers.

A slide containing the experimenter's instructions was presented first. The instructions were as follows: "You are going to solve a series of 5-letter anagrams shown on the screen. They will appear only for a short time. Work the anagram out in your head and write the answer in the space provided. Numbers below each anagram will help you to keep track." This was followed by a practice session in which five example anagrams were presented. After this practice session a participants were asked if they had any questions, and any issues arising were clarified. The full set of eighty anagrams was then presented to the class. After the last anagram, a final slide was shown confirming the end of the study and thanking participants for their efforts.

Analysis

Rasch analysis allows us to investigate how well a variable, in this case anagram solution score, meets the criterion of being both unidimensional and reliable and also creates interval level data of solution difficulty. There are many Rasch models, but data resulting from a dichotomous outcome are governed by a probabilistic process for the linear combination of two parameters, one denoting person ability and the other denoting item difficulty.

The basic model is:

$\text{Log} (P_{ni1} / P_{ni0}) \equiv B_n - D_i$ where

B_n is the ability of subject *n*, where *n* = 1, N.

D_i is the difficulty of item *i*, where *i* = 1, L.

P_{ni1} is the probability of subject *n* succeeding on item *i*.

P_{ni0} is the probability of failure $1 - P_{ni1}$.

This is expressed as

$$P_{ni1} = e^{(Bn - Di)} / 1 + e^{(Bn - Di)} .$$

Its application to the analysis of anagram solution difficulty was facilitated using *WINSTEPS* (Linacre, 2005).

The data matrix from this study converged rapidly with only three *PROX* passes and four *UCON* passes. The Prox method (Cohen, 1979) is used to get rough estimates of the Rasch measures for both persons and items. These estimates are then used by UCON (unconditional maximum likelihood estimation; Wright and Panchapakesan, 1969) which fine tunes them through iteration to produce the final estimates.

Rasch item separation (see Wright & Stone 1979, 1996) was 5.08 with an item reliability of 0.96. Person separation was 3.62 with a person reliability of 0.93. These outcomes suggest a well-designated and indexed variable responded to in a cogent manner by the subjects. The reported item reliability is equivalent to the familiar *KR-20* or *Cronbach's* α. The high value of 0.96 for items indicates that a cohesive variable has been conceived based upon a working theoretical strategy for how subjects would respond. The person reliability of 0.93 is almost as high. This statistic is less familiar in the literature of test development, but it is no less important (Wright & Stone 1996). The high value suggests that the variable is being addressed by most respondents as intended.

Regression analysis will also be used to investigate the calibration of the difficulty of the anagrams by good and poor solvers. This is a useful technique to look at the possibility of differential item functioning, in this case that the anagrams are not being solved differently by the two groups.

Results

Each anagram was given a solution score (a possible 0-128) equal to the number of participants who solved it. Solution scores were reliable, as there was an inter group correlation of r (78) = .931, p <.005 between the two testing sessions. There was no significant relationship between the position of each anagram in the list and its solubility (r (80) = .036, p = .75).

Rasch scaling, using *WINSTEPS* (Linacre, 2005), produces a scaling map of items and persons (see Figure 1). This map lists the items and persons on the same variable scale from the most difficult items at the top of the table downward to the less difficult items. On the same scale, the person ability distribution ranges from most able person at the top downward to less able persons in descending order. The scaling procedure centred the mean of the 80 items at 50 logits with a standard deviation of 12.10.

The importance of number of syllables on the difficulty of solution was investigated using these estimates. As can be seen in Figure 1, twenty-four of the thirty-nine items above the mean were multi-syllable items, whereas of the forty-one below the mean, only seven were multi-syllable. In total, there were two three-syllable items, twenty-nine two-syllable items, and forty-nine one syllable items. Thirty-six of the one-syllable items were at or below the mean of item difficulty. The mean difficulty of multi-syllable words (56.78) was significantly higher than that of one-syllable words (45.7; t (79) = 4.4, p <.001). Number of syllables is significantly correlated with Rasch score (r (80) = .446, p <.0005). Hence, there is clear evidence that one-syllable anagrams are easier to solve than multi-syllable anagrams.

There was one anagram word which was an outlier in these analyses, and that is SCYTH, a one-syllable word that has in fact the highest Rasch score (see Figure 1). This is an unusual word as it contains no vowels and also does not appear in many frequency counts. Furthermore, its spelling is problematic as several dictionaries identify this as a six letter word SCYTHE, so it is omitted from the remaining analyses.

There was also some support for the view that syllable frequency (Macizo and Van Patten, 2007) affects solution difficulty. There was a significant correlation between the log of first syllable frequency and anagram Rasch score (r (79) = .409, p<.0005), but this became non significant when number of syllables was controlled (r (76) =.104, p>.05). However, if only the 31 multisyllabic words were examined, there was a significant correlation between the log of first syllable frequency and anagram Rasch score (r (31) =.386, p<.05). This means that the more frequent the syllable then the more difficult it was to solve the anagram, presumably because there are more competing similar words if the syllable is more frequent.

Following Gihooly and Johnson (1978) a stepwise regression analysis was carried out including the twelve variables which were used by them and also including the three word frequency measures mentioned earlier and number of syllables. The variable which was selected first was mean pronounceability with a multiple R of .538 and an adjusted R^2 of .279. This replicates previous research which has shown a pronounceability effect (Dominowski, 1969; Gilhooly & Johnson, 1978; Novick & Sherman, 2008). The pronounceability effect is partly a feature of the letters in a word, which are a property of the word, but also partly of random organisation. It is possible to make an anagram of a word more or less pronounceable, which is why Novick and Sherman (2008) refer to pronounceability as a superficial feature of anagram solution and one which will not be useful in diagnosing how problems are solved.

Accordingly, another regression analysis was conducted in which pronounceability was entered first, followed by a stepwise procedure to select from the other independent variables. Number of syllables was the first variable to be entered after pronounceability which raised the multiple R to .633 and the adjusted R^2 to .4. The only other variables to be entered were GTZERO and log the Kucera-Francis frequency (see Table 1). The importance of GTZERO has been noted before (Mendelsohn 1976). The regression equation with all four variables entered had a multiple R of .711 and an adjusted multiple R^2 of .47

Although it has been argued that pronounceability is a superficial characteristic of an anagram (Novick and Sherman, 2008) and not useful in determining how problems are solved, it is important to note that to some extent pronounceability will depend on the letters involved in the word from which the anagram is formed. Accordingly by including it in the regression first we might be diminishing the importance of the other variables structural variables which are correlated with it. Specifically, the number of vowels (r (79) =.494, p<.0005), GTZERO (r (79) =.471, p<.0005) and the number of syllables (r (79) =.408, p<.0005) are all significantly correlated with the pronounceability of the anagram and hence might have their influence diminished by including this as the first variable in the regression. To counteract this, these variables were regressed onto pronounceability and the residuals were saved as a variable which would represent the effects of pronounceability without the effects of these variables. The multiple R for this equation was .571 which indicated how much of the effect of the structural variables will be mistakenly included in the superficial variable of pronounceability.

The multiple regression analysis was performed again using the residual of the regression onto pronounceability as the first variable to be entered and then allowing stepwise selection from all of the variables. The results are presented in Table 2. They were very similar to the previous results except that GTZERO now became the most important variable with number of syllables slightly less important than before. The Kucera-Francis frequency variable has also been replaced with a measure of word familiarity but these are highly correlated (r (79) =.616, p<.0005). Both analyses support the idea that the number of syllables and GTZERO are important variables in determining anagram solution. A new variable was created to investigate the interaction of GTZERO and syllables by multiplying them together and found that this variable has a very high correlation with Rasch solution score (r (79) =.633, p<.0005). Furthermore, when only the 31 multi-syllable words were examined, the residual for pronounceability and GTZERO were the only variables entered into a stepwise regression equation with a multiple R of .63.

Novick and Sherman (2008) found that their good anagram solvers were more likely to be affected by structural features such as the number of syllables than poor anagram solvers. This was examined directly using the Rasch analysis. The total sample of 128 subjects was utilized for computing all the statistics reported above. A subsequent division of this sample was made into two groups: the sixty-four highest measured persons and the sixty-four lowest. Each sample was separately calibrated with *WINSTEPS* to produce high-group and low-group item difficulties. These values are plotted in Figure 2.

A linear regression line is indicated in the figure with $R^2 = 0.82$. A polynomial equation of third degree produced a curvi-linear fit of $R^2 = 0.92$. The high fit to both groups is another indication that the variable is cohesive and that high and low measured persons are responding in a similar fashion to the items.

Discussion

The results of the present study support Novick and Sherman's (2008) finding that syllable number is a major factor in determining anagram difficulty, at least for five-letter words. The more syllables the target word contained the harder it was to solve and the higher the Rasch score. One possible explanation for this is that multiple syllables provide greater competition in terms of the possible letter combinations that have to be tried in order to find the correct word. It appears that anagram solvers use syllable structure as an initial guide in regrouping the letters in the anagram. As Goldblum and Frost (1988) concluded, "… syllables appear to be the best cue for word retrieval" (p.164). It is also clear that previous research which has not included syllable number as a variable needs to be reassessed.

However, we do not claim that syllable number is the sole contributor to anagram difficulty. It is clear that syllable number is a confounding factor for other variables which are likely to contribute. These include starting letter and syllable frequency. Anagrams of words that begin with a vowel are more difficult to solve, but they are more likely to contain more than one-syllable. In the current study words randomly chosen that began with a vowel included 'ankle', 'index' and 'orbit'. Similarly, syllable frequency may well make a contribution to anagram difficulty as it does in naming tasks (e.g. Conrad & Jacobs, 2004), however, in the current study it was clear that more frequent syllables were confounded with syllable number and the effect of frequency was only visible in multisyllabic words.

Novick and Sherman (2008) argued that the syllable structure of the anagram is particularly important to good anagram solvers. Our results suggest that syllable number has an impact on all solvers, or at least on all undergraduate solvers. It is, of course, likely that Novick and Sherman's good solvers were more skilful than our solvers and that an even more pronounced effect would be shown by this group of participants. However, it is important to point out that we replicate Novick and Sherman's (2008) syllable results with a new larger sample of non-selected participants, with a new selection of stimuli and robust measurement.

Rasch measurement takes into account two task parameters; item difficulty and the ability of the participant, plotting both of them on a unidimensional scale. Although these parameters are to some extent interdependent, Rasch modelling separates the parameters by using a probabilistic approach in which a participant's raw score on the test is converted into a success-to-failure ratio. Hence, the probability of a participant being successful on a given item is a log function of the difference between that participant's ability and the difficulty of the item. This approach has advantages over traditional measurement in that the scales produced are genuine interval measures, rather than assumed interval measures. They are, therefore, appropriate for statistical analysis based on interval scales. Rasch analysis also allows us to use information about the ability of the participants without conducting a pre-test, which in some cases will have an affect on the results of the experiment itself. However, most importantly it goes some way to meeting the criteria of conjoint measurement (Luce and Tukey, 1964; Perline, Wright, and Wainer, 1979), rather than just assigning numbers according to a rule (Stevens, 1946). It is increasingly recognised that Rasch measurement is important in the development of useful psychometric scales, but it is equally important that experimental measures have the properties which are fundamental to real measurement.

Overall, we suggest this research makes a useful contribution to measurement models of human cognitive problem-solving. The results presented need to be extended to other word lengths but it is clear that syllables play an important role in accessing words in our mental lexicon. It is also clear that when using anagrams in experimental studies, care needs to be taken in selecting anagrams to ensure that differences in syllable number do not confound the results.

References

Balota, D. A., Cortese, M.J., Hutchison, K.A., Loftis, B., Neely, J.H., Nelson, D., Simpson, G.B., & Treiman, R. (2002). *The English Lexicon Project: A web-based repository of descriptive and behavioral measures for 40,481 English Words and Nonwords*, Washington University, http://elexicon.wustl.edu/

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133*, 283-316.

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition, 29*, 639–647.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database (CD-ROM)*, Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

Beversdorf, D. Q.,Ferguson, J. L. W., Hillier, A.,, Sharma, U. K , Nagaraja, H. N., Bornstein, R. A., & Scharre, D. W. (2007). Problem Solving Ability in Patients With Mild Cognitive Impairment. *Cognitive Behavioral Neurology, 20*, 44–47.

Clark, H.H. (1973). The language as fixed-effect fallacy: A critique of language statistics in psychological research. Journal of Verbal Learning and Verbal Behaviour, 12, 335-339.

Cohen, L. (1979) Approximate expressions for parameter estimates in the Rasch model. *British Journal of Mathematical Psychology, 32*, 113-120.

Coleman, E.B. (1964). Generalizing to a language population. *Psychological Reports*, 14, 219-226.

Conrad, M., & Jacobs, A. M. (2004). Replicating syllable frequency effects in Spanish in German: One more challenge to computational models of visual word recognition. *Language and Cognitive Processes, 19*, 369-390.

Dominowski, R. L. (1966). Anagram solving as a function of letter moves. *Journal of Verbal Learning and Verbal Behavior, 5*, 107-111.

Endler, N. S., Speer, R. L., Johnson, J. M., & Flett, G. L. (2001). General self-efficacy and control in relation to anxiety and cognitive performance. *Current Psychology, 20 (1)*, 36-52.

Foley, M. A., & Foley, H. J. (2007). Source-monitoring judgments about anagrams and their solutions: Evidence for the role of cognitive operations information in memory. *Memory & Cognition, 35 (2)*, 211-221.

Gilhooly, K., & Hay, D. (1977). Imagery, concreteness, age-of-acquisition, familiarity, and meaningfulness values for 205 five-letter words having single solution anagrams. *Behavior Research Methods & Instrumentation, 9 (1)*, 12-17.

Gilhooly, K., & Johnson, C.E. (1978). Effects of solution word attributes on anagram difficulty: A regression analysis. *Quarterly Journal of Experimental Psychology, 30*, 57-70.

Goldblum, N., & Frost, R. (1988). The crossword puzzle paradigm: The effectiveness of different word fragments as cues for the retrieval of words. *Memory and Cognition, 16 (2)*, 158-166.

Kucera, H., & Francis, W. (1967). *Computational analysis of present day American English*. Providence, R.I.: Brown University Press.

Linacre, J. (2005). *WINSTEPS*. Chicago: MESA Press.

Luce, R. D., & Tukey, J. E. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology, 1(1)*, 1-27.

Macizo, P., & Van Petten. (2007). Syllable frequency in lexical decision and naming of English words. *Reading & Writing, 20 (4)*, 295-331.

Mendelsohn, G. A. (1976). An hypothesis approach to the solution of anagrams. *Memory & Cognition, 4*, 637-642.

Novick, L. R., & Sherman, S. J. (2003). On the nature of insight solutions: Evidence from skill differences in anagram solution. *Quarterly Journal of Experimental Psychology, 56 (2)*, 351-382.

Novick, L. R., & Sherman, S. J. (2008). The effects of superficial and structural information on on-line problem solving for good versus poor anagram solvers. *Quarterly Journal of Experimental Psychology, 61*, 1098-1120.

Perline, R., Wright, B., & Wainer, H. (1979). The Rasch model as additive conjoint measurement. *Applied Psychological Measurement, 3 (2)*, 237-255.

Ronning, R.R. (1965). Anagram solution times: A function of the “rule-out” factor. *Journal of Experimental Psychology, 69*, 35-39.

Stenneken, P., Conrad., & Jacobs, A. M. (2007). Processing of syllables in production and recognition tasks. *Journal of Psycholinguistic Research, 36*, 65-78.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*, 677-680.

Thorndike, E. L., & Lorge, I. (1944). *The teacher's word book of 30,000 words*. New York: Teacher's College Press.

Wright, B.D. & Panchapakesan, N (1969). A Procedure for Sample Free Item Analysis, *Educational and Psychological Measurement (EPM), 29*. 23-48.

Wright, B., & Stone, M. (1979). *Best test design*. Chicago: MESA Press.

Wright, B., & Stone, M. (1996). *Measurement essentials*. Wilmington, DE: Wide Range, Inc.

Table 1. *Results for the multiple regression analysis for the 79 solution words.*

Multiple R	.71***
Pronounceability	β= ..29**
Number of syllables	β= .48**
GTZERO	β= .33**
Log of Kucera Francis	β= -.22*

Stepwise selection after pronounceability entered and presented in order of entry.

(*)*p* <.10. **p* <.05. ***p* <.01. ****p* <.001.

Table 2. *Results for the multiple regression analysis with the residual of pronounceability entered first.*

Multiple R	.73***
Residual for pronounceability	β= .24**
GTZERO	β= .41***
Number of syllables	β= .29**
Familiarity	β= -.23**
Similarity	β= -.20*

Stepwise selection after residual for pronounceability entered and presented in order of entry.
**p* <.05; ***p* <.01; ****p* <.001.