
Recebido: 01-03-2023 | Aprovado: 15-03-2023 | DOI: <https://doi.org/10.23882/rmd.23136>

Speed and Power Indices of tests and items

Índices de velocidade e potência de testes e itens

Satyendra Chakrabarty, Indian Statistical Institute, Indian Maritime University, Indian Ports Association (chakrabartysatyendra3139@gmail.com)

Abstract: A simple index is proposed to measure speed or power components of a test. The index is independent of position of the items and provides necessary and sufficient condition for pure speed test and pure power test and enables testing of statistical hypothesis to infer that the test can be taken as a speed test or a power test. Similar index of an item is also proposed to reflect whether the item is a speed item or a power item. The proposed index C is a ratio such that $C = 0 \Leftrightarrow$ Pure power test and $C = 1 \Leftrightarrow$ Pure speed test facilitating computation of similar index of each item and statistical test of significance. Properties of the index discussed. Operational method outlined to modify a test to speed or power test. Items can be ranked with respect to such item-wise index. Identification of power items and speed items help to modify the test to a speed or power test by deleting items in stages, if speediness (or power) is not intended. Relationship between index for the test and item-wise indices derived.

Keywords: Error scores, Unattempted items, Random guessing, Speed test, Power test.

Resumo: Um índice simples é proposto para medir componentes de velocidade ou potência de um teste. O índice é independente da posição dos itens e fornece condição necessária e suficiente para teste de velocidade puro e teste de potência puro e permite o teste de hipótese estatística para inferir que o teste pode ser feito como um teste de velocidade ou um teste de potência. Índice semelhante de um item também é proposto para refletir se o item é um item de velocidade ou um item de potência. O índice C proposto é uma razão tal que $C = 0 \Leftrightarrow$ *Pure power test* e $C = 1 \Leftrightarrow$ *Pure speed test* facilitando o cálculo de índice semelhante de cada item e teste estatístico de significância. Propriedades do índice discutidas: método operacional descrito para modificar um teste para teste de velocidade ou potência. Os itens podem ser classificados em relação a esse índice de itens. A identificação de itens de potência e itens de velocidade ajuda a modificar o teste para um teste de velocidade ou potência, excluindo itens em etapas, se a velocidade (ou potência) não for pretendida. Relação entre o índice para o teste e os índices de itens derivados.

Palavras-chave: Pontuações de erro, Itens não tentados, Suposição aleatória, Teste de velocidade, Teste de potência.

1. Introduction:

Major challenges of tests relate to assessment of “ability” for power tests and “speed” for speed tests. However, ability and speed jointly affect response behavior in tests (Partchev et al., 2013; Van der Linden, 2009). Primary sources of individual differences in speed tests and power tests are speed of response or speed of information processing (SIP) and accuracy of response. Abilities measured by a test under speeded conditions are different than the same measured under un-speeded conditions (Lord, 1956). Van der Linden (2009) found low or even negative correlations between accuracy and response time across persons. Speed may be manifested in the form of random guessing, number of unattempted items, inattentiveness, etc. Subjects taking a test may need to adjust between accuracy and time to maximize his/her score. Speeded responses do not depend solely on a test taker's ability and are therefore not appropriate for traditional item response theory (Cintron, 2021). However, an inattentive response is broader than a pure random response (Meade & Craig, 2012).

Methods of analysis like item analysis, reliability, validity, etc. and interpretation of scores are different for these two types of tests. For example, split-half reliabilities are erroneously high for speed tests and may be taken as an upper bound for the reliability coefficient (Gulliksen, 1950). Substantive degrees of speededness tend to underestimate validity of tests (Lu and Sireci, 2007). Reliability and validity of speed tests are influenced by the speediness component, since variance of speed test is not due to the mental ability of interest. Problems get aggravated since most tests are combination of unknown proportion of speed and power, which makes development of appropriate theorems in test theory more difficult than for pure type tests (Gulliksen, 1950).

Thus, question arises on quantification of speed and power components of a test and the items. Constructs of ability and speed are common primarily in cognitive domains. However, response times are also considered in non-cognitive domains like personality, attitudes, etc. (Ferrando & Lorenzo-Seva, 2007; Ranger & Kuhn, 2012). Attempts have been made to isolate the speed component which is not related to the

level of interest in speeded tests using external information like response times and also not using any external information (see Lu & Sireci, 2007). Based on the Stafford's Speediness Quotient (SQ, 1971) for items, Estrada et al. (2017) separated speed and power components of tests of mental ability without considering other information like response times where a rule of thumb was suggested for identifying items affected by speediness.

Need is felt to derive measures reflecting degree of speed and power of a test. The paper proposes an index C as a ratio such that $C = 0 \Leftrightarrow$ Pure power test and $C = 1 \Leftrightarrow$ Pure speed test facilitating computation of similar index of each item and statistical test of significance. Properties of the index discussed. Operational method outlined to modify a test to speed or power test.

2. Literature survey:

2.1 Definitions and Important terms:

In a speed test, items are so easy that if a subject attempts an item, he/she gets it correct. However, due to large number of items and insufficient time, nobody can finish the test within the specified time limit. Time limit of a power test is chosen so that each subject gets opportunity to attempt all the items. But some items are so difficult that all subjects cannot give correct answer to each and every item of the test. Thus, in a speed test, score differences reflect variations in speed of response and in a power test; score differences indicate variations in accuracy of responses.

Different types of error scores in the context of speed –power issues are:

Wrong answers (W) refer to the items for which a subject failed to answer correctly. Unattempted items (U) are the number of omitted items (a subject decided not to answer after reading those items – primarily for power test) and Not-reached items (not attempted due to less availability of time – primarily for speed test).

Thus, error score E is given by

$$E = W + U \quad (1)$$

It may be noted that subjects can answer the items in any sequence say from the end or by skipping alternate items, etc. Thus, it is not justified to consider Not-reached items as those at the end of a test. For practical purpose, unattempted items are items not endorsed by the subjects.

2.2 Measures of Speed and Power:

Attempts made to measure speededness by two administrations of a test with and without time limits. Using two such administrations, Cronbach and Warrington (1951) suggested a measure denoted as *tau* (τ) in terms of correlations between test scores, corrected for attenuation. However, tau does not consider difference of scores under speed and power administrations. Strictly speaking, two versions of the test under speed and power may not be parallel since mean and variance are likely to vary for the two versions. In other words, if Version 1($v1$) and Version 2($v2$) are parallel, at least the following two conditions need to be satisfied:

- Mean of $(X_{v1} - X_{v2}) = 0 \Rightarrow \overline{X_{v1}} = \overline{X_{v2}}$
- $Var(X_{v1}) = Var(X_{v2})$

From single administration of a test, and denoting standard deviation of U-scores, W-scores and error scores respectively by S_U , S_W and S_E , Gulliksen (1950) proposed the following two inequalities:

$$\text{For power tests: } 1 + \frac{S_U}{S_E} > \frac{S_W}{S_E} > 1 - \frac{S_U}{S_E} \quad (2)$$

$$\text{For Speed test: } 1 + \frac{S_W}{S_E} > \frac{S_U}{S_E} > 1 - \frac{S_W}{S_E} \quad (3)$$

However, Rindler, (1979) showed difficulty in interpretations of contribution of speed as per Gulliksen's inequalities when $\frac{S_U}{S_E}$ or $\frac{S_W}{S_E}$ are large. Focusing on proportion of total errors due to speededness, Stafford (1971) estimated Speededness Quotient (SQ), as $SQ = \frac{\sum U}{\sum W + \sum U}$ or percentage of unattempted items in the error scores. It allows for assessment of speed both on individual level and on test level. Swineford (1974) suggested that in an essentially unspeeded test, at least 80% of the examinees reach the

last item and every examinee reaches at least 75% of the items. However, this “arbitrary criterion” is not rigidly applied (College Entrance Examination Board (CEEB), 1984).

IRT based approaches involving set of assumptions like unidimensionality, local independency etc. have been adopted to estimate speededness from single-administration of tests. Hambleton et al. (1991) considered 3-PL IRT model defined as

$$P_i(\theta) = C_i + (1 - C_i) \frac{e^{a_i(\theta - b_i)}}{1 + e^{a_i(\theta - b_i)}} \text{ where}$$

$P_i(\theta)$: Probability of answering a random item correctly by a subject with ability θ

a_i : item discrimination

b_i : item difficulty value and

C_i : pseudo guessing parameter

Bejar (1985) proposed an item-level index and an examinee-level index, making further assumptions. But values of both the indices may vary depending on other sources of error confounded with the effect of speededness and interpretations of the indices are difficult (Lu & Sireci, 2007).

Effect of random guessing due to speededness has given contrasting results. With small amount of random guessing due to speededness, Attali (2005) found largely attenuate inter-item correlations, and attenuate Cronbach's alpha, but large amount of random guessing due to low-motivation could result in inflated Cronbach's alpha (Wise & DeMars, 2009). Major reason of different conclusions could be use of real data by the later and conclusions based on analytical derivations and simulations by the former. Other factors, like pooled samples, can potentially inflate reliability (Flinn et al., 2015). Random responses are independent of item content and the latent trait of the respondent and may arise due to speededness, low motivation, inattentiveness, and tendencies of respondents to rush to maximize attempted number of items.

Models for response times differ in approaches, assumptions, statistical distributions considered, complexities and findings. Different statistical distributions were used in different models viz. Log normal (Van der Linden, 2007), Gamma (Maris, 1993), Weibull (Rouder et al., 2003), Box-Cox transformation to approximate almost any distribution (Klein et al., 2009), etc.

Impact of speeded responses on item-total correlations and Cronbach's alpha were studied by Hong and Cheng (2019) with two types of manifestations of test speediness, i.e. random guessing versus reduced ability and found that that inter-item correlations may inflate or deflate in different cases depending on the combinations of item parameters, the mean Cronbach's alpha rarely increases under simulations using real test parameters, even with different manifestations of speededness. Thus, inflated Cronbach's alpha may be an artifact of a sample and not a population behavior. However, other manifestations of speededness giving rise to insufficient effort responding (IER) to survey are there (Huang et al., 2012). Despite the issue of inflated or deflated inter-item correlations, factor analysis of SAT data was undertaken (CEEB, 1984) and found that factors attributable to speed accounted for about 5% to 10% of the variance of test scores.

IRT with flexibility in choosing data collection plan offers important advantages. However, conceptually and procedurally complex IRT is based on strong assumptions, satisfaction of which need to be tested. For example, IRT assumes that the probability of an examinee to answer an item correctly does not depend on whether the item is placed at the beginning, in the middle, or at the end of the test. Probability of hitting the correct answer by guessing only cannot be determined by usual IRT model.

3. Proposed method:

For the i -th subject, let E_i be the total error score which is sum of W_i (number of wrong answers) and U_i (number of un-attempted items i.e. non-reached items + omitted items). From equation (1),

$$E_i = W_i + U_i$$

If the test consisting of K -number of items is administered to n -subjects, one can have mean of error score is equal to sum of mean of W-scores and U-scores i.e.

$$\bar{E} = \bar{W} + \bar{U} \quad (4)$$

$$\text{and } S_E^2 = S_W^2 + S_U^2 + 2r_{wu}S_W S_U \quad (5)$$

An index C ($0 \leq C \leq 1$) is conceptualized to measure degree of power and degree of speed as $C = \frac{\bar{U}}{U_{Max}} = \frac{1}{nK} \sum U_i$ (6)

where U_{Max} is obtained when everybody fails to attempt even a single item i.e. $U_i = K \forall i = 1, 2, \dots, n$. Thus, $C = \frac{\bar{U}}{K}$ is a ratio lying between zero and one for a general test.

For a pure power test, $U_i = 0 \forall i = 1, 2, \dots, n$. This implies $C = 0$ for a pure power test. Conversely, $C = 0 \Rightarrow \sum U_i = 0 \Rightarrow U_i = 0 \forall i = 1, 2, \dots, n$ i.e. the test is a pure power test.

Following similar logic, it can be proved that $C = 1$ for a pure speed test and $C = 1 \Rightarrow \bar{U} = K \Rightarrow U_i = K \forall i = 1, 2, \dots, n$ i.e. the test is a pure speed test. Thus, pure power test $\Leftrightarrow C = 0$ and pure speed test $\Leftrightarrow C = 1$. In other words, necessary and sufficient condition for a pure power test is $C = 0$ and the same for a pure speed test is $C = 1$. In practice, one may not always get a power test for which $C = 0$ and can make statistical test to see whether the obtained value of C is significantly different from zero i.e. testing $H_0: \bar{U} = 0$. Alternately, the obtained value of $C > 0$ may be taken as a measure of departure from the pure power position.

3.1 Pure Power test:

A pure power test can be formally defined as follows:

Definition 1: A test X is said to be a pure power test if and only of the index C_x as defined in (6) is zero for the test.

For all practical purposes, a test X can be considered as a power test if index C_x is not significantly different from zero. Rejection of $H_0: \bar{U} = 0$ implies that the test cannot be regarded as a pure power test.

The proposed index helps to improve the criterion of power test given by Gulliksen (1950) by using the following theorem:

Theorem 1. Let Y_1, Y_2, \dots, Y_n be n -independent observations of a variable Y such that $Y_i \geq 0 \forall i = 1, 2, \dots, n$. If \bar{Y} is closed to zero ($\bar{Y} \approx 0$), then $S_Y^2 \approx 0$

Proof: If each $Y_i = 0$, the theorem is trivially true. Assume all Y_i 's are not equal to zero.

Call $\bar{Y} = \epsilon$ where ϵ is a small positive number.

$$\text{Then, } \bar{Y}^2 = \frac{1}{n^2} (Y_1 + Y_2 + \dots + Y_n)^2 = \epsilon^2$$

$$\Rightarrow \frac{1}{n^2} \sum_{i=1}^n Y_i^2 + \frac{2}{n^2} \sum_{i \neq j} Y_i Y_j = \epsilon^2$$

$$\Rightarrow \frac{1}{n^2} \sum_{i=1}^n Y_i^2 < \epsilon^2 \text{ since } \frac{2}{n^2} \sum_{i \neq j} Y_i Y_j > 0$$

$$\Rightarrow \frac{1}{n} \sum_{i=1}^n Y_i^2 < n\epsilon^2$$

$$\Rightarrow \frac{1}{n} \sum_{i=1}^n Y_i^2 - \bar{Y}^2 < n\epsilon^2 - \epsilon^2$$

$$\Rightarrow S_Y^2 < (n-1)\epsilon^2$$

$$\Rightarrow S_Y^2 \approx 0 \text{ since } \epsilon \approx 0$$

Remarks: Converse of the theorem is not true since if each observation is equal to a large number (say ∇), then $S_Y^2 = 0$ but $\bar{Y} = \nabla$

3.2 Improving Gulliksen's criteria for power test:

$$\text{Gulliksen's criterion for power test is } 1 + \frac{S_U}{S_E} > \frac{S_W}{S_E} > 1 - \frac{S_U}{S_E} \quad (7)$$

$$\text{For a pure power test } \bar{U} = 0 \Rightarrow \frac{\bar{U}}{K} = 0 \Rightarrow C = 0$$

$$\text{As per the theorem 1, } C = 0 \Rightarrow S_U = 0 \Rightarrow \frac{S_U}{S_E} = 0$$

$$\text{Thus, for } C = 0, \frac{S_W}{S_E} = 1; 1 + \frac{S_U}{S_E} = 1 \text{ and } 1 - \frac{S_U}{S_E} = 1 \text{ and the inequality (7)}$$

becomes

$$1 + \frac{S_U}{S_E} \geq \frac{S_W}{S_E} \geq 1 - \frac{S_U}{S_E} \quad (8)$$

However, converse is not true i.e. $\frac{S_U}{S_E} = 0$ does not imply $U = 0$. Consider a test where $U_i = M \forall i = 1, 2, \dots, n$ and M is large positive integer less than or equal to the total number of items. Here, $S_U = 0 \Rightarrow \frac{S_U}{S_E} = 0$ but $\bar{U} = M$ and the test is not a power test.

So, $C = 0 \Leftrightarrow$ Pure power test is more general statement than Gulliksen's criterion $\frac{S_U}{S_E} = 0$.

In fact, $C = 0$ is the necessary and sufficient condition for a pure power test. If a test is moderately power, then $C \approx 0 \Rightarrow \bar{U} \approx 0 \Rightarrow S_U \approx 0$ by the theorem which implies $\frac{S_U}{S_E} \approx 0$ and inequality (2) holds.

3.3 Index of Speed test:

For a pure speed test, the index $C = 1$ and vice versa. When number of unattempted items for each subject is equal to total number of items in the test, $C = 1$. One can test $H_0: C = 1$ and $(1 - C)$ can be taken as departure from pure speed test. So, a pure speed test is defined as follows:

Definition 1.2: A test X is said to be a pure speed test if and only of the index C_x as defined in (6) is equal to one for the test.

Gulliksen's condition for speed test (inequality 3) can be improved considering $C = 1$. From equation (6), $C = 1 \Rightarrow \bar{U} = K$

Thus, $\bar{E} - \bar{U} = K - \bar{E}$ since $\bar{E} \geq \bar{U}$ and $K \geq \bar{E}$.

In other words, $\bar{E} = K$ if $C = 1$

As per equation (4), $\bar{E} = \bar{W} + \bar{U}$. Putting $\bar{E} = K$ for $C = 1$, for a pure speed test, $\bar{E} = \bar{U} \Rightarrow \bar{W} = 0$. Using Theorem 1, we have $S_W^2 = 0 \Rightarrow \frac{S_W}{S_E} = 0$

For a pure speed test, $W_i = 0 \Rightarrow \bar{W} = 0$ and $S_W = 0$. From (1), $S_U = S_E$

Thus, $1 + \frac{S_W}{S_E} = 1$; $\frac{S_U}{S_E} = 1$; $1 - \frac{S_W}{S_E} = 1$

Accordingly, Gulliksen's criterion for speed test (3) boils down to

$$1 + \frac{S_W}{S_E} \geq \frac{S_U}{S_E} \geq 1 - \frac{S_W}{S_E} \quad (9)$$

Therefore, for $C = 1$, Gulliksen's condition for speed test is improved to accommodate pure speed test. However, the converse is not true. Thus, Gulliksen's condition is true for one way only. $C = 1 \Leftrightarrow$ Pure speed test is more general statement and is a necessary and sufficient condition for a pure Speed test.

3.4 Speed and Power items:

Consider the matrix U with n -rows for n -examinees and K -columns for K -number of items, where the $(i-j)$ th cell = 1 if the i -th individual has not attempted the j -th item and

= 0 if the i -th individual has attempted the j -th item

Here, total of j -th column gives total number of unattempted items by the sample of examinees.

The C -index for the j -th item $C_j = \frac{\text{Total of } j\text{-th column}}{\text{Number of examinees}} = \frac{\sum_{i=1}^n U_{ij}}{n} = \bar{U}_j$ (10)

Clearly, maximum value of $C_j = 1$ when no examinee could attempt the j -th item indicating that the j -th item is a pure speed item. Minimum value of $C_j = 0$ when each examinee attempted the j -th item indicating that the j -th item is a pure power item.

The items may be ranked with respect to C_j and thus facilitate identification of speed items along with assessment of degree of speededness. In reality, C_j may be closed to one ($C_j \approx 1$) $\forall j = 1, 2, \dots, K$.

Thumb rule of accepting j -th item as a speed item if $C_j \geq 0.5$ is arbitrary. Better will be to undertake testing of $H_0: C_j = 1$. Acceptance of $H_0: C_j = 1$ imply the j -th item is a speed item and rejection of $H_0: C_j = 1$ indicates that the j -th item is not a speed item. Similar exercise can be undertaken for power items with $C_j \approx 0$ along with identification of power items.

If C -index of the test is denoted by C_{Test} ,

$$\text{average of } C_j's = \frac{1}{K} \sum_{j=1}^K \bar{U}_j = \frac{1}{nK} \sum U_i = C_{Test} \quad (11)$$

The equation (11) gives relationship between C_{Test} and C_j 's

Identification of power items and speed items help to modify the test to a speed or power test by deleting items in stages.

4. Discussion:

The proposed index of speediness looks similar to the Speededness Quotient (SQ) proposed by Stafford (1971). SQ is defined as the percentage of unattempted items in the total number of errors for individual level and on test level. $SQ = 100$ for a purely speeded test and for a purely power test $SQ = 0$. Like the proposed C -index, SQ focuses on proportion of total errors unlike the proportion of test variance affected by speed in Gulliksen's approach. The proposed index in terms of $(E_i - W_i)$ for the i -th individual may not have a one-to-one correspondence with SQ . Tests with no penalty for wrong answers will significantly decrease value of SQ . For example, about 99.6% of the examinees answered each item in the Swedish Scholastic Aptitude Test (SweSAT) (Marcus, 2021) for which SQ will be closed to zero irrespective of speededness or power components of SweSAT. In addition, the C -index helps one to test $H_0: C = 1$ and helps to identify items measuring speed.

5. Limitations:

The proposed indices of test and items cannot help to find effect of random guessing with different manifestations of speededness. Values of the indices are also affected by homogeneity or heterogeneity of the sample.

6. Conclusion:

A simple index C_{Test} ($0 \leq C_{Test} \leq 1$) in terms of a ratio is proposed for measuring the degree of speed or degree of power of a test. The index C_{Test} is independent of the position of the items and is equal to $\frac{\bar{U}}{K}$ where \bar{U} denotes the mean number of unattempted items by n -examinees under a prescribed time limit. The test becomes close to a power test as C_{Test} tends to zero and close to a speed test as C_{Test} tends to 1. Converse is also true. In fact, necessary and sufficient condition for a pure power test is $C_{Test} = 0$ and the same for a pure speed test is $C_{Test} = 1$. Gulliksen's inequalities separately for power test and speed test modified to include pure power test and pure speed test.

The index C_{Test} facilitates statistical test to see whether the obtained value of C_{Test} is significantly different from zero i.e. testing $H_0: \bar{U} = 0$. In case of rejection of the null hypothesis, $(1 - C_{Test})$ may be taken as a measure of departure from the pure power position. Similarly, for speed test, one can test $H_0: C_{Test} = 1$ and $(1 - C_{Test})$ can be taken as departure from pure speed test.

Following similar approach, C -index for the j -th item C_j was defined, which reflects a pure power item if $C_j = 0$ and a pure speed item if $C_j = 1$. The items of the test can be ranked with respect to C_j and help in identification of speed items along with assessment of degree of speededness. In reality, C_j may be closed to one ($C_j \approx 1$). Acceptance of statistical hypothesis $H_0: C_j = 1$ implies that the j -th item is a speed item and rejection of $H_0: C_j = 1$ indicates that the j -th item is not a speed item. Similar exercise can be undertaken for power items with $C_j \approx 0$ along with identification of power items. Identification of power items and speed items help to modify the test to a speed or power test by deleting items in stages, if speediness (or power) is not intended. Relationship between C_{Test} and C_j 's derived. The method can be best applied when it is desirable to minimize test speededness or when speed is not a part of the latent trait being measured.

Future empirical studies with real and/or simulated data sets may be undertaken for further investigation of the indices and effect on psychometric qualities.

References:

- Attali, Y. (2005). Reliability of speeded number-right multiple-choice tests. *Applied Psychological Measurement, 29*, 357-368. <https://doi.org/10.1177/0146621605276676>
- Bejar, I. I. (1985). *Test speededness under number-right scoring: An analysis of the Test of English as a Foreign Language*. Report No. ETS-RR-85-11, Educational Testing Service. Princeton, NJ
- Cintron, D. (2021). Methods for Measuring Speededness: Chronology, Classification, and Ensuing Research and Development. *ETS Research Report Series, 2021(1)*. <https://doi.org/10.1002/ets2.12337>
- College Entrance Examination Board (1984). *The College Board technical handbook for the scholastic aptitude test and achievement tests*. New York
- Cronbach, L. J., & Warrington, W. G. (1951). Time limit tests: Estimating their reliability and degree of speeding. *Psychometrika, 14*, 167-188.
- Estrada, E., Román, F. J., Abad, F. J., & Colom, R. (2017). Separating power and speed components of standardized intelligence measures. *Intelligence, 61*, 159-168.
- Ferrando, P. J., & Lorenzo-Seva, U. (2007). An item response theory model for incorporating response time data in binary personality items. *Applied Psychological Measurement, 31(6)*, 525-543. <https://doi.org/10.1177/0146621606295197>
- Flinn, L., Braham, L., & Das Nair, R. (2015). How reliable are case formulations? A systematic literature review. *British Journal of Clinical Psychology, 54*, 266-290. <https://doi.org/10.1111/bjc.12073>
- Gulliksen, H. (1950). *Theory of Mental Tests*. New York, Wiley, 177 - 198
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory* (Vol. 2). SAGE Publications, Inc. Newbury Park. California 91320
- Hong, M. R., and Cheng, Y. (2019). Clarifying the Effect of Test Speededness. *Applied Psychological Measurement, 43(8)*, 611-623. <https://doi.org/10.1177/0146621618817783>
- Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012): Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology, 27*, 99-114. <https://doi.org/10.1007/s10869-011-9231-8>
- Klein Entink, R. H., van der Linden, W. J., & Fox, J. P. (2009): A Box-Cox normal model for response times. *British Journal of Mathematical and Statistical Psychology, 62*, 621-640.
- Lord, F. (1956). A study of speed factors in tests and academic grades. *Psychometrika, 21*, 31-50.

- Lu, Y., & Sireci, S. G. (2007). Validity issues in test speededness. *Educational Measurement: Issues and Practice*, 26(4), 29–37. <https://doi.org/10.1111/j.1745-3992.2007.00106.x>
- Marcus S. Hjärne (2021). Just Enough Time to Level the Playing Field: Time Adaptation in a College Admission Test, *Scandinavian Journal of Educational Research*, 65(6), 941-955. <https://doi.org/10.1080/00313831.2020.1788143>
- Maris, E. (1993). Additive and multiplicative models for gamma distributed random variables, and their application as psychometric models for response times. *Psychometrika*, 58, 445-469.
- Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, 17, 437-455. <https://doi.org/10.1037/a0028085>
- Partchev, I., De Boeck, P., & Steyer, R. (2013). How much power and speed is measured in this test? *Assessment*, 20(2), 242–252. <https://doi.org/10.1177/1073191111411658>
- Ranger, J., & Kuhn, J.T. (2012). Improving item response theory model calibration by considering response times in psychological tests. *Applied Psychological Measurement*, 36(3), 214–231. <https://doi.org/10.1177/0146621612439796>
- Rindler, S. E. (1979). Pitfalls in assessing test speededness. *Journal of Educational Measurement*, 16(4), 261–270.
- Rouder, J., Sun, D., Speckman, P., Lu, J., & Zhou, D. (2003). A hierarchical Bayesian statistical framework for response time distributions. *Psychometrika*, 68, 589-606.
- Stafford, R. E. (1971). The speed quotient: A new descriptive statistic for tests. *Journal of Educational Measurement*, 8, 275–278.
- Swineford, F. (1974). *The test analysis manual* (SR-74-06). Princeton, NJ: Educational Testing Service.
- Van der Linden, W. J. (2009). Conceptual issues in response-time modeling. *Journal of Educational Measurement*, 46(3), 247–272. <https://doi.org/10.1111/j.1745-3984.2009.00080.x>
- Van der Linden, W. J. (2007). A hierarchical framework for modeling speed and accuracy on test items. *Psychometrika*, 72, 287-308.
- Wise, S. L., & DeMars, C. E. (2009). A clarification of the effects of rapid guessing on coefficient a: A note on Attali's reliability of speeded number-right multiple-choice tests. *Applied Psychological Measurement*, 33, 488-490. <https://doi.org/10.1177/0146621607304655>