

UNIVERSITY OF CAMBRIDGE, UNITED KINGDOM,
DEPARTMENT OF PSYCHIATRY¹,
CHARLES UNIVERSITY, CZECH REPUBLIC,
DEPARTMENT OF KINANTHROPOLOGY²,
UNIVERSITY OF YORK, UNITED KINGDOM,
DEPARTMENT OF HEALTH SCIENCES AND GYMS³

ON THE MEASUREMENT LEVEL OF HANDEDNESS SCORES

JAN ŠTOCHL^{1,2}, TIM CROUDACE³

SUMMARY

The scoring of handedness items is generally performed using a scale of “always use the right hand” to “always use the left hand”. Although there is no guarantee that such scores are *ordinal* measures, the data from these tests are traditionally analysed with statistical methods that assume ordinality. This study therefore aims to examine the measurement level of handedness data and specifically focuses on verification of the ordinality. The results suggest that the categorical items used in handedness questionnaires are indeed ordinal.

Key words: handedness, level of measurement, item response theory, Bock’s nominal response model

INTRODUCTION

An item is the basic unit of any test or scale. Items can be statements, commands or questions, the answers (responses) to which are associated in clear unequivocal terms with the desired trait/ability to be measured. An item often allows for a series of response options, which are known as “selected-response” items. Alternatively, an item may require a respondent to construct his/her own responses, which are known as “constructed-response” items. Whereas multiple-choice, true-false, short-answer, and essay item formats are commonly used in educational measurement practice, the rating scale format is the most popular format for psychological measurements. In laterality research, the scales for the assessment of hand preference also generally employ a rating scale format, with response categories ranging from *always use the right hand* to *always use the left hand*.

All of the above-mentioned formats can be assigned to a level of measurement. Although various conceptualisations of measurement levels have been developed, four levels are generally recognised in the methodological literature, following the definitions of Stevens: nominal (also known as unordered categorical), ordinal (ordered categorical), interval and ratio scales of measurement. The key distinguishing feature between nominal

and ordinal scales is that the assigned numbers represent the rank order of the entities assessed by the ordinal scale. Interval scales have all of the properties of ordinal scales, and any differences between the levels of the attributes can be multiplied by any real number to exceed or equal other inter-level differences. The “zero” point rating is arbitrary on interval scales but not on ratio scales. Even today, there is a continuing debate regarding this classification.

Scores of various types of assessment, e.g., personality measures, such as Eysenck’s Personality Questionnaire or the Portrait Value Questionnaire, are scored on scales that have two extremes. In the EPQ, for example, the scales are neuroticism/stability and extraversion/introversion. Hand preference scales have a similar property, ranging from extreme right hand preference to extreme left hand preference. The level of the measurement of scores (nominal or ordinal) depends on the wording of the responses. If researchers use scoring items such as *never use the right hand*, *rarely use the right hand*, *occasionally use the right hand*, *often use the right hand*, and *always use the right hand*, then the scale would generally be considered ordinal. For the purposes of a handedness assessment, the wording of the response categories mentioned above is inconvenient because researchers should inquire about the same activity for the left hand. It is therefore more convenient to use a combined wording for both hands, i.e., *always use the right hand*, *prefer to use the right hand*, *use both hands equally often*, *prefer to use the left hand*, and *always use the left hand*. This bipolar wording is common in handedness studies, although the number of response categories varies by study. These wording and scoring methods have been proposed by White & Ashton as alternatives to those of the Edinburgh Handedness Inventory. This revision has been used in several studies.

For statistical analyses, it is common practice to assign numerical values to each response category, for instance, 1 = *always use the right hand*, 2 = *prefer to use the right hand*, ... , and 5 = *always use the left hand*. Given the original wording of response categories, however, it is clear that such data do not have a ratio level of measurement. Moreover, the interval scale is also unlikely to be appropriate, and even ordinal levels of measurement may not be presumed. The level of measurement of the data is crucial for decisions regarding the most appropriate analytic tools to use for data processing. Parametric statistical methods cannot be used for nominal data and have a somewhat limited usability for ordinal data.

The aim of this study was to assess the measurement level of data from a hand preference questionnaire that uses bipolar wording. The results of these analyses may encourage the use of suitable statistical methods in future research that makes use hand preference scales. The secondary aim of the essay was to provide a discussion on the optimal number of response categories to use in measurement tools for handedness.

SAMPLE AND METHODS

Sample and measures

The dataset analysed here was composed of the responses of 2,055 individuals who completed an online questionnaire. Of these, 529 (25.7%) were excluded due to

incomplete or apparently invalid responses, e.g., random responses or responses with no variability. The analysed sample comprised the remaining 1,526 (868 females, 584 males, and 74 unknown) respondents from 97 countries. The mean age of respondents was 33.0 years (sd = 12.8 years). Given the online nature of the research and voluntary participation of the respondents, no consent was signed by the respondents.

An online questionnaire (available at <https://www.surveymonkey.com/s/T8LXXSZ>) was developed in 3 languages (English, Czech, and Chinese). This computer-based survey included 4 items that measured handedness (*holding a toothbrush, hammering a nail, throwing a ball, and unlocking a door*); These items were chosen because they are appropriate for individuals of all known cultures (unlike, for example, writing). Furthermore, these items make sense within the context of several adaptive theories on the origin of laterality. These four handedness items are used here to evaluate the level of measurement of hand preference scores.

The handedness responses were recorded on 5-point scales, with the responses arranged in the following manner: always use the right hand = 1, prefer to use the right hand = 2, use both hands equally often = 3, prefer to use the left hand = 4, and always use the left hand = 5.

Analysis

The ordinality or nominality of the scores was examined using Bock's nominal response model. The NRM is a very flexible model in that certain other IRT models, such as the partial credit model or generalised partial credit model can be considered to be constrained versions of NRM. The distinguishing feature of the NRM from the above-mentioned IRT models is that the item responses are assumed to be unordered (i.e., nominal). However, the ordinality of the response categories can be empirically tested, as described in the following paragraphs.

Let X_i denotes the score on item i . Let us suppose this item has m response categories. The item characteristic curve (ICC) of NRM can then be written as follows:

$$P_{ui}(\theta) = P(X_i = u | \theta) = \frac{e^{(a_u\theta + c_u)}}{\sum_{k=0}^{m-1} e^{(a_k\theta + c_k)}}$$

for $u = 1, \dots, m_i - 1$ and $k = 1, \dots, m_i - 1$, and where θ denotes the latent trait value. The parameters a_{ik} , c_{ik} are referred to as the slope and intercept, respectively, and characterise the shape of the individual ICC. This model is not fully identified; therefore, an additional constraint must be imposed:

$$\sum_{k=0}^{m-1} a_{ik} = \sum_{k=0}^{m-1} c_{ik} = 0$$

a_{ik} and c_{ik} are related to discrimination and difficulty, respectively. Then, the NRM can be reformulated as follows:

$$P_{ui}(\theta) = \frac{e^{(a_{iu}\theta + c_{iu})}}{\sum_{k=0}^{m-1} e^{a_{ik}(\theta - b_{ik})}}, \text{ where } b_{iu} = \frac{a_{iu}}{c_{iu}}.$$

The parameters a_{iu} and b_{iu} can therefore be interpreted as the slope (discrimination) and location (difficulty), respectively.

Under the condition of a good fit of the NRM model to the data, Samejima Demonstrated that if the slopes a_{iu} monotonically increase along the response categories, then the response categories are ordinal.

All IRT models, including the NRM, require “essential” unidimensionality of the items. However, it is not a simple matter to evaluate if this assumption holds if the level of measurement is unknown. Use of factor analysis would be problematic because it assumes at least ordinality. Therefore, we must rely on face validity and content validity, which was assessed and found to be relevant by experts during the development and initial validation of the questionnaire.

To assess the ordinality of the data, the NRM was fitted and the response category slopes were checked for monotonicity. The fit of the NRM was evaluated by comparing the observed proportions in each response category to the proportion predicted by the model.

Subsequently, we determined the ICCs and discussed the optimal number of response categories for each item.

RESULTS

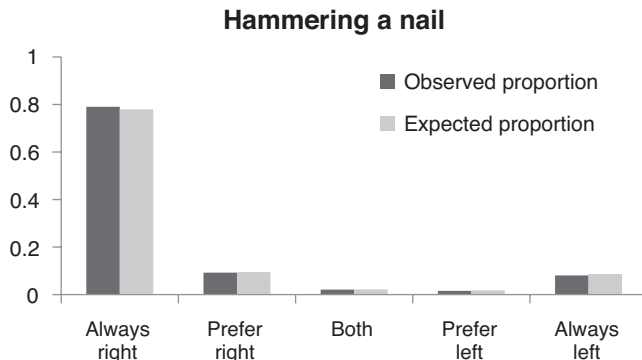
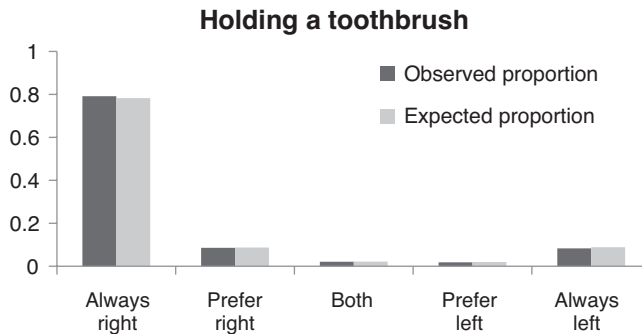
Evaluation of the ordinality of the response categories

Table 1 lists the slopes for each response category after fitting the data to the NRM. For all of the items, these slopes were monotonically increasing. Therefore, the response categories can be considered to be ordinal.

Table 1. Slopes within the Bock’s nominal response model

Item	Response category				
	Always right hand	Prefer right hand	Both hands equally often	Prefer left hand	Always left hand
Throwing a ball	-4.48	-3.28	-1.61	3.34	6.03
Hammering a nail	-651	-3.14	-1.91	3.23	8.34
Unlocking a door	-2.57	-1.12	-0.54	1.46	2.76
Holding a toothbrush	-3.93	-1.89	-0.87	1.98	4.71

The figures in the previous table are trustworthy only in the case that the NRM fits the data. This fit can be expressed as a comparison of the observed proportions of responses in each response category with the proportions predicted by the model. The next figure illustrates these two quantities. The values in this figure (Figure 1) indicate a close fit of the NRM to the data.



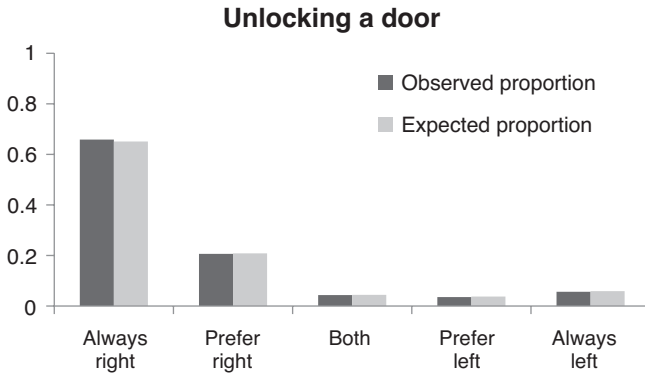


Figure 1. Fit of the Bock's nominal response model to the data

Optimal number of response categories

Figure 2 illustrates the item characteristic curves for the NRM. These curves indicate the probability of responding with the response categories as a measure of handedness. Handedness is assumed to lie on a continuum (the x-axis), from extreme right-handedness (this category is on the lower values on x-axis due to the direction of item scoring) to extreme left-handedness (higher values on x-axis).

In the other paper (Štochl, Croudace, 2012, in preparation), we find the interpretation of handedness as continuous variable to be somewhat limited and argue instead for a categorical nature of handedness. For the purposes of this article, however, it is convenient to consider handedness as a continuous variable.

It can be observed that for all of the items, the response category “use both hands equally often” is less useful because the probability of answering these responses is lower than the probability of answering other response categories over the entire latent continuum. Similarly, the response category “prefer to use the left hand” is not the most probable response category for any level of handedness for any of the items, except for *throwing a ball*. Generally, the number of “useful” response categories varies from two (for the item *holding a toothbrush*) to four (for *throwing a ball*).

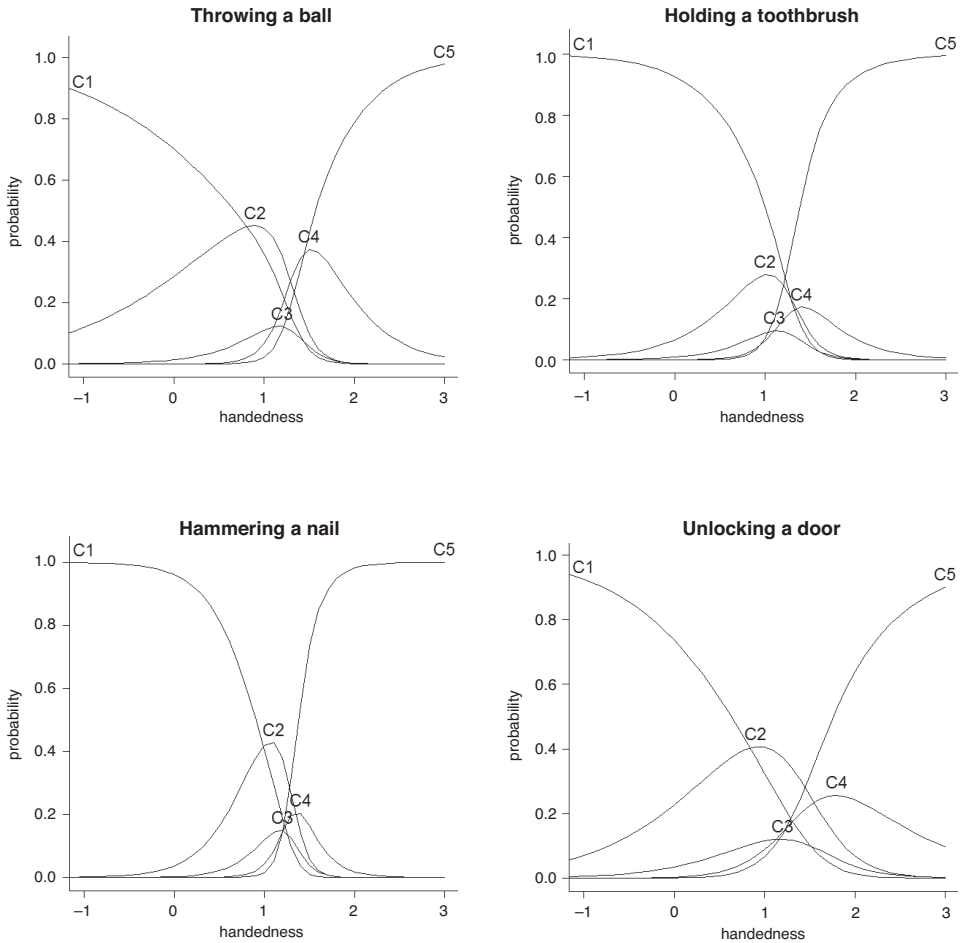


Figure 2. Estimated item characteristic curves using the Bock's nominal response model
 Solid Lines: Always use the right hand = C1; prefer to use the right hand = C2; use both hands equally often = C3; prefer to use the left hand = C4; and always use the left hand = C5.

DISCUSSION

Researchers who examine handedness often use questionnaires scored on bipolar scales ranging from *always use the right hand* to *always use the left hand*. This study presents an investigation of the measurement level of hand preference responses to such questionnaires. Four handedness-related items from an online questionnaire were used for this purpose. The item response modelling was employed to statistically test the ordinality of the response categories.

The most general IRT model, the NRM, was fit to the data. The ordinality of the data was tested by estimating the slopes of the model. The data can be considered ordinal if

the slopes within the item are monotonically increasing. In our case, this was true for all 4 of the handedness items under consideration. These results therefore suggest that our data can be considered as ordinal.

The assessment of implicit ordering is important because the analytical methods used in the following essays assume the ordinality of the items. This is also true for nearly all of the analytical methods commonly used for data processing in laterality research. Generally, the ordinality of data from handedness questionnaires is implicitly assumed. However, to our knowledge, this assumption has rarely been empirically tested. In certain cases, the ordinality of the response categories is obvious and does not require verification (for example, for Likert-type scales). However, this is not a case for laterality research, as the bipolar wording of response categories does not assure ordinality. Researchers who ignore this important issue by assigning numbers to their response categories cannot be certain that higher values indicate “more” of the measured trait, or that lower numbers signify “less”. In this sense, nominal scales are qualitative rather than quantitative. It should now be clear that performing any statistics (except for determining the mode) with these numbers makes little sense.

Apart from response category ordering, this study also briefly discusses the number of response categories that are optimal or “useful” for handedness items. We found that the category “use both hands equally often” is less useful because conditional probability of this response is lower than for the other responses, regardless of hand preference. This is to certain extent also true for the category “prefer to use the left hand”. However, suggesting the avoidance of these categories is not a simple matter, primarily because the decrease in the number of response categories is generally believed to decrease the amount of obtained information. For example, the effect of the number of response categories on the magnitude of the Pearson correlation coefficient is well known; generally, this coefficient tends to be lower as the number of response categories decreases. Other studies have also demonstrated that a decrease in the number of categories may decrease the internal consistency, test-retest reliability, item validity, and inter-rater reliability. In addition, as noted by Van der Linden & Hambleton, the joint probability of the less-chosen response categories for a particular level of the measured trait may be higher than the probability that any other single response category is chosen.

Alternatively, fewer categories do not necessarily lead to a low internal consistency statistic. A study by Chang, for example, found that more scale categories do not necessarily enhance reliability and may even lead to a systematic “abuse” of the scale. Thus, to determine the impact of categorisation on the internal structure of a measure is required before a suggestion is made regarding the optimal number of categories for handedness items. However, such an extensive assessment was outside of the scope of this study and is left for future research.

The analyses presented in this study can (to a certain extent) be viewed as ex-post justification of the results of previous laterality studies in which similar wording of response categories was used. However, the results from this study require cross-validation in different populations and for different types of scoring of hand preference responses before a sound generalisation to other handedness questionnaire can be made.

ACKNOWLEDGEMENTS

This study was supported by the Czech Science Foundation, grant nr. 406/07/P513, by the Czech Ministry of Education (research project MSM 0021620864) and by British National Institute for Health Research programme grant RP-PG-0606-1335.

REFERENCES

- ADAMS, E. W. (1966). On the nature and purpose of measurement. *Synthese*, 16(2), pp. 125–169.
- BANDALOS, D. L. & ENDERS, C. K. (1996). The effects of nonnormality and number of response categories on reliability. *Applied Measurement in Education*, 9(2), pp. 151–160.
- BOCK, R. D. (1972). Estimating item parameters and latent ability when the responses are scored in two or more nominal categories. *Psychometrika*, 37, pp. 29–51.
- BRYDEN, M. P. (1977). Measuring handedness with questionnaires. *Neuropsychologia*, 15(4–5), pp. 617–624.
- CHANG, L. (1994). A psychometric evaluation of 4-point and 6-point likert-type scales in relation to reliability and validity. *Applied Psychological Measurement*, 18(3), pp. 205–215.
- CICCHETTI, D. V., SHOINRALTER, D., & TYRER, P. J. (1985). The effect of number of rating scale categories on levels of interrater reliability: A Monte Carlo investigation. *Applied Psychological Measurement*, 9(1), pp. 31–36.
- COHEN, J. (1983). The cost of dichotomization. *Applied Psychological Measurement*, 7, pp. 249–253.
- DUNCAN, O. D. (1984). *Notes on social measurement: Historical and critical*. New York: Russell Sage Foundation.
- EYSENCK, H. J., & EYSENCK, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire*. London: Hodder and Stoughton.
- LUCE, R. D., KRANTZ, D. H., SUPPES, P., & TVERSKY, A. (1990). *Foundations of measurement: Vol. 3. Representation, axiomatization, and invariance*. San Diego, CA: Academic.
- MARTIN, W. S. (1973). The effects of scaling on the correlation coefficient: A test of validity. *Journal of Marketing Research*, 10, pp. 316–318.
- MARTIN, W. S. (1978). Effects of scaling on correlation coefficients: Additional considerations. *Journal of Marketing Research*, 15, pp. 304–308.
- MASTERS, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, pp. 149–174.
- MASTERS, J. R. (1974). The relationship between number of response categories and reliability of Likert-type questionnaires. *Journal of Educational Measurement*, 11(1), pp. 49–53.
- MESSINGER, H. B., MESSINGER, M. I., & GRAHAM, J. R. (1988). Migraine and left-handedness: Is there a connection? *Cephalalgia*, 8(4), pp. 237–244.
- MICHELL, J. (1997). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology*, 88(3), pp. 355–383.
- MURAKI, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16, pp. 159–176.
- OLDFIELD, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9(1), pp. 97–113.
- PRESTON, C. C., & COLMAN, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104(1), pp. 1–15.
- RAYMOND, M., & PONTIER, D. (2004). Is there geographical variation in human handedness? *Laterality*, 9(1), pp. 35–51.
- SAMEJIMA, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometric Monograph no 17*.
- SCHACHTER, S. C., RANSIL, B. J., & GESCHWIND, N. (1987). Associations of handedness with hair color and learning disabilities. *Neuropsychologia*, 25(1B), pp. 269–276.
- SCHWARTZ, S. (1994). Are there universal aspects in the structure and contents of human values? *Journal of Social Issues*, 50(4), pp. 19–45.

- SHERMAN, J. L., & KULHAVY, R. W. (1976). *The assessment of cerebral laterality: The Sherman-Kulhavy laterality assessment inventory. Technical report no. 4.* Arizona State University.
- STEVENS, S. S. (1946). On the theory of scales of measurement. *Science*, 103, pp. 677–680.
- THISSEN, D. & STEINBERG, L. (1986). A taxonomy of item response models. *Psychometrika*, 51(4), pp. 567–577.
- VAN DER LINDEN, W. J., & HAMBLETON, R. K. (eds). (1997). *Handbook of modern item response theory.* New York: Springer.
- WHITE, K., & ASHTON, R. (1976). Handedness assessment inventory. *Neuropsychologia*, 14(2), pp. 261–264.

METODA EMPIRICKÉ VERIFIKACE ORDINALITY DAT NA PŘÍKLADU DOTAZNÍKU LATERALITY

JAN ŠTOCHL, TIM CROUDACE

SOUHRN

Článek představuje metodu empirické verifikace ordinality dat pomocí nominálního modelu z rodiny modelů teorie položkových odpovědí. Ukázka použití metody je na datech dotazníku laterality od 1526 respondentů z 97 zemí světa. Výsledky studie ukazují, že data z položek rukové preference v tomto dotazníku mohou být považována za ordinální.

Klíčová slova: lateralita, škálování, teorie položkových odpovědí, Bockův nominální model

Jan Štochl, PhDr., Ph.D., MPhil
stochl@ftvs.cuni.cz

Tim J Croudace, BSc., DipAppPsych, MSc., Ph.D.
tim.croudace@york.ac.uk