Neuropsychologia, Vol. 28, No. 5, pp. 443–455, 1990.
Printed in Great Britain.

0028-3932/90 $3.00+0.00

TACHISTOSCOPIC PRESENTATION OF VERBAL STIMULI FOR ASSESSING CEREBRAL DOMINANCE: RELIABILITY DATA AND SOME PRACTICAL RECOMMENDATIONS

MARC BRYSBAERT and GÉRY D'YDEWALLE

Department of Psychology, University of Leuven, B-3000, Leuven, Belgium

(*Received* 27 *April* 1989; *accepted* 13 *November* 1989)

Abstract—Reliability data point to rather high test–retest correlations (≥ 0.65) for VHF data with four- and five-letter words as stimuli, but replicate previous findings that the first test score correlates poorly with later test scores. The same results are obtained for accuracy and latency data, though small differences exist. All laterality indices lead to the same conclusions and have high intercorrelations. The point-biserial correlation coefficient is, however, a slightly more reliable index of naming latency than the mere difference between LVF and RVF. No such superiority is found for the indices based on accuracy data. The results also point to the need to present a sufficient number of stimuli before firm conclusions can be drawn.

INTRODUCTION

DESPITE thirty years of intensive research it still is rather difficult to set up a visual recognition task for determining cerebral dominance, a variable that is of interest in clinical as well as in experimental settings. One reason for this is the lack of reliability data, as stressed by MCKEEVER [15] in 1986 and still true. Another reason is that most studies involve a comparison of groups, providing little data about individual performances and therefore little information about which standards to use for assessment.

With respect to the first problem, FENNELL *et al.* [8] were the first to investigate the reliability of accuracy data in a visual half-field (VHF) task. They assessed the test–retest stability of a sequential letter pairs task, in which four different letter pairs were presented sequentially, each pair exposed for 300 msec. One letter of each pair appeared at the fixation point and the other appeared in the Left Visual Field (LVF) or Right Visual Field (RVF). Subjects reported all letters at fixation before the lateralized letters. Subjects were tested weekly over a 4-week period. It was found that LVF- and RVF-score had a significant test–retest reliability between days 2 and 3, 2 and 4, and 3 and 4 (Pearson correlations ranging from 0.55 to 0.86). No significant correlation was found between the first day scores and later day scores. A similar analysis of left-handed subjects showed comparable but slightly lower reliability data [9, 12, but see 7]. Grant [10] assessed test–retest reliability for the recognition of four-letter words in 11-year-old children. Subjects were tested twice within 3 days. Pearson correlations of 0.14 and 0.82 were obtained for poor and good readers respectively. Performance reliability also differed between boys and girls. ROSZKOWSKI and SNELBECKER [19] investigated the temporal stability of the chimeric face technique in 14-year-old children for two test sessions 1 month apart. A stable VHF superiority was found in 67% of the subjects. The typical LVF bias was more reliable than the unusual RVF bias.

MCKEEVER [15], finally, reported split-half correlations of LVF minus RVF latencies (msec) for his Object Naming Latency task and Colour Naming Latency task. A coefficient of 0.48 was found for Object Naming and 0.76 for Colour Naming. Ninety-six percent of the subjects showed a consistent half-field superiority over the halves in the Colour Naming Latency test, against 74% for the Object Naming Latency test.

The second problem for setting up a VHF task that assesses cerebral dominance in individuals has to do with the lack of standards in current literature. At first, things seem rather simple: subjects who show an atypical visual field superiority (e.g. LVF superiority for verbal stimuli, RVF superiority for clock reading or face recognition) are classified as right cerebral dominant, the others as left cerebral dominant. However, a closer look reveals some major difficulties with this strategy. First, VHF tasks generally yield an estimate of about 30% right hemisphere dominance in right-handed subjects, whereas more invasive techniques as the Wada test yield estimates of only about 3% [23]. It has been argued that the difference may be due to the fact that invasive and non-invasive assessments of lateralization address different processing mechanisms [2], but a conclusive proof for this statement has yet to be given. Second, there may be a difference in the number of subjects showing an atypical visual field superiority according to the task administered. Studies with words typically show a greater RVF advantage than studies with other stimuli, and studies with long words, a greater RVF advantage than studies with short words [27]. There has been a long debate about whether these differences are due to a difference in hemispheric specialization or to other factors such as reading habits, directional scanning or visual acuity [5, 11, 13, 21, 27, 28]. Third, as the reliability studies mentioned above indicate, there are rather low test–retest correlations in VHF tasks. This means that subjects showing an atypical superiority pattern in one session need not show the same pattern in subsequent sessions, leaving us with the question what criterion to use. Fourth, tachistoscopic studies result in two dependent variables: response accuracy and reaction latency. Which one is to be used, or can they provide converging evidence, as BABKOFF and FAUST [1] hold?

In what follows, we will mainly be concerned with the reliability of a particular VHF task for determining cerebral dominance using words as stimuli. The question of standards for assessment will be dealt with elsewhere [6]. As will become clear, the VHF task can only yield reliable results if one is willing to invest a fair amount of energy. Efforts have been made to use equipment available in most experimental and clinical settings. As it turned out, the minimal requirement only is a microcomputer with text mode processing and two "software screens".

METHOD

Subjects

Fourteen male subjects participated on a voluntary basis. Three of them were the first author (right-handed) and two research assistants (left-handed); the other subjects were psychology undergraduate students. Half of the subjects had strong right hand, right foot, right ear, and right eye preferences as measured with PORAC and COREN's questionnaire [18], the other half had strong left preferences for the same activities.

Stimuli

The stimulus pool consisted of a sample of 100 four-letter words (Dutch) and 100 five-letter words. Nouns, verbs, adjectives and function words were mixed to get a sample of the total language corpus [25]. The four-letter words consisted of three types: (a) high frequency words (mean frequency = 1335/270 000), (b) low frequency words that only differed from the high frequency words by the first letter (f = 15/270 000), and (c) low frequency words that only differed from the high frequency words by the last letter (f = 11/270 000). The five-letter words consisted of higher frequency (f = 303/270 000) and lower frequency words (f = 55/270 000). The main aim of the frequency manipulation was to see whether word-beginning and word-end have a different weight in the LVF and the RVF.

This is important because one of the criticisms against VHF studies with words is that the RVF is favoured for more reasons than just laterality. It is said that the beginning of a word contains more information than the end, and that, therefore, the RVF is additionally favoured because in the RVF the word-beginning lies nearer to the fixation location than in the LVF [5, 11, 13, 21, 27, 28]. Therefore, words that are prone to be misguessed on the basis of either the first or the last letter were included. This manipulation was not possible for the five-letter words. Word frequency *per se* was not expected to have a significant effect, because it failed to affect the results in prior studies [16, 27] and the words in this study all were presented more than once (see the following).

Apparatus and procedure

Stimulus presention was controlled by an IBM XT microcomputer, programmed to present a random sample from the four- or the five-letter word pool left of fixation and the remaining words right of fixation. The advantage of the procedure was that no prediction could be made as to which side a particular word would appear. This was important because all words were presented five times. Two disadvantages of the procedure were: (a) that the number of stimuli presented on one side changed slightly from one session to the other, though still remaining close to 50% left and 50% right, and (b) that the number of times a specific word was presented in the left or right field could vary between 0 and 5. The first disadvantage was corrected by taking the proportion of correct answers instead of the absolute number. The second disadvantage will be coped within the Results section. Chances of having zero presentations on one side were 2/32.

Stimuli were presented in lower case letters for 140 msec in such a way that all RVF words (four-letter words as well as five-letter words) began six letter positions from fixation location, and all LVF words ended six letter positions from fixation location. The four-letter words had a length of 12.2 mm, the five-letter words a length of 15.3 mm. Subjects sat about 40 cm away from the screen (there were no head restraints), so that stimulus width was 1.75° for the four-letter words, and 2.20° for the five-letter words. The distance between fixation location and beginning of a word in RVF or end of a word in LVF also was 2.20°. Screen persistence of the CRT display was circumvented by displaying a one-sided mask that consisted of 4 or 5 ASCII codes 178 (#) immediately after the stimulus. Stimulus luminance amounted to 24 cd/m^2 (measured with a Minolta NT-1 for the mask) against a screen of 3 cd/m^2. The mask remained until the subject finished the response.

Series of four- and five-letter words were alternated. Four right-handed and three left-handed subjects started with a four-letter series, the remaining with a five-letter series. Stimuli were randomly presented to the left and to the right of fixation to minimize expectancy and hemispace effects [3]. A trial consisted of a delay period randomly ranging from 2000 to 3000 msec, a warning signal of 150 msec, a second delay period of 750 msec, stimulus presentation for 140 msec, and mask presentation. The fixation mark (ASCII code 92, '_') remained visible throughout. Subjects were instructed to report as rapidly as possible the word that had been presented. Naming latency was registered with the use of a voice trigger connected to the game port. If the subjects did not recognize the word, they were encouraged to guess or to report the letters they had seen. The experimenter keyed in the response (visible for the subject on the screen) and asked whether he had understood the answer correctly. Thereafter, mask and answer disappeared and a new trial began.

In order to further ensure that the subjects really fixated the fixation mark, after a random number of trials (varying between 0 and 9) a random digit instead of a new stimulus was presented above the fixation mark for 60 msec, followed by a mask (ASCII 178). Subjects had to name the digit. If they made a mistake, they were warned by a tone. The experimenter told them that more than 10% of errors made the series invalid. It took the subjects on an average two series before the criterion of less than 10% was reached. Though there was no digit to be named on each trial, the strategy made the subjects very alert to look at the fixation mark on all trials.

A series of 100 words and about 20 digits took about 20 min to complete. At the beginning of the experiment, subjects were told how to use the microphone and what stimuli would be presented. They then were given two sample trials. After each series, subjects got feedback about their mean reaction time for the correct answers, the number of mistakes they had made, and the number of data which had to be dropped out because of "technical problems". The latter was an index of outliers (see Results). Subjects finished five replications of the four- and the five-letter series. This took about 4 hr, divided among two or three sessions (depending on the time the subjects could free themselves from courses and other activities) within a period of at least 3 days and at most 1 week.

RESULTS

Accuracy data

A $2 \times 2 \times 2 \times 5$ ANOVA including one between (Handedness) and three within-subjects (VHF, Wordlength, and Series) variables was calculated on the proportion correct answers. The variables Handedness, Wordlength and VHF were fixed, the variable Series random. The quasi F-ratios needed for the analysis were determined according to Cochran and calculated with the use of a program by WOLACH and MCHALE [26].

Three main effects were significant: VHF (LVF = 0.48, RVF = 0.76; $F(1, 12) = 54.40$, $P < 0.0001$), Wordlength (four-letter words = 0.67, five-letter words = 0.58; $F(1, 11) = 33.2$, $P < 0.001$), and Series (first series = 0.55, second = 0.62, third = 0.62, fourth = 0.67, and fifth = 0.67; $F(4, 48) = 16.45$, $P < 0.0001$). A significant Wordlength × VHF interaction emerged because the accuracy of the five-letter words was lower than the accuracy of the four-letter words in the LVF and equal in the RVF (four-LVF = 0.56, five-LVF = 0.41, four-RVF = 0.77, five-RVF = 0.76; $F(1, 13) = 11.62$, $P < 0.01$). This interaction changed slightly during the different replications giving rise to a significant Wordlength × VHF × Series interaction ($F(4, 48) = 2.85$, $P < 0.05$). Overall accuracy ranged from 0.45 to 0.81 over subjects.

Percentage correct responses are usually transformed in a so called "laterality index" in order to assess cerebral asymmetry. The simplest of these indices is the difference (d) between the RVF and LVF score. However, because the d index is quite vulnerable to ceiling and floor effects, a number of "better" indices have been proposed. They include (a) the difference score to the corrected total (d_0), (b) the percentage of correct responses (POC), (c) the percentage of errors (POE), (d) the e index (e), (e) the phi coefficient (phi) and (f) the lambda index (L). All the indices are listed in Table 1. For a more formal account of the indices and their interrelations, see BRYDEN [4] and SPROTT and BRYDEN [24]. Because d_0 and POC are linear transformations of each other [4], only d_0 will be considered further on.

Table 1. The different laterality indices based on accuracy data and their computation. P_R stands for the probability of a correct response in the RVF, P_L for the probability of a correct response in the LVF, R+ and L+ for the number of correct responses in the RVF and LVF, and R− and L− for the number of wrong responses in RVF and LVF

d	$P_R - P_L$
d_0	$(P_R - P_L)/(\frac{1}{2}(P_R + P_L))$
POC	$P_R/(P_R + P_L)$
POE	$(1 - P_L)/(2 - P_R - P_L)$
e	$(P_R - P_L)(P_R + P_L)$ if $(P_R + P_L) \leq 1$
	$(P_R - P_L)/(2 - P_R - P_L)$ if $(P_R + P_L) > 1$
Phi	$(P_R - P_L)/\{(P_R + P_L)(2 - P_R - P_L)\}^{1/2}$
L	$\ln(R+/R-) - \ln(L+/L-)$

Before analysing what the individual indices reveal, it might be interesting to look at their intercorrelations over subjects. In what follows, intercorrelations always are Spearman's *rho* coefficients [22]. We prefer this coefficient because it is linearly related to Kendall's *W* coefficient which will be used further on (see SIEGEL [22]), and it involves fewer assumptions than Pearson's correlation coefficient [22].*

Table 2 displays the correlations between the mean indices for the four- and the five-letter words. They range from 0.53 between d_0 and POE to 1.00 between POE and e, which indicates

*As a matter of fact, all analyses have been done with three correlation coefficients: Spearman's *rho*, Kendall's *tau*, and Pearson's r. As a rule Kendall's *tau* was slightly more conservative than Spearman's *rho*, and Pearson's r slightly less conservative.

that all indices point to the same pattern of results. Correlations may be lower, if the data are not centered around 60% accuracy but are more extreme, because the different indices have been suggested in particular to remedy floor and ceiling effects [4, 24].

ANOVAs with three factors (Handedness-fixed, Wordlength-fixed, and Series-random) yielded a significant main effect of Wordlength ($P<0.01$) for all indices, and a significant interaction between Wordlength and Series for *d*, d_0, and *phi* ($P<0.05$). Further examination, however, showed that the interaction was more a matter of small fluctuations and not of any significant difference in trend.

Of greater importance is the test–retest reliability of the different indices. Because of space economy and clarity, we will not present full intercorrelation matrices for every index, but limit ourselves to three measures per index (see Table 3): a Kendall *W* reliability coefficient for the four- and the five-letter words, and a rank correlation for the coherence between the four- and the five-letter words. Kendall *W* coefficients [22] were calculated for the four-letter and the five-letter words separately because of the significant difference between them. The *W* coefficient bears a linear relation to the average *rho* taken over all possible pairs of rankings.

Four things are to be noted in Table 3. First, all *W*'s for the four-letter words lie between 0.52 and 0.60. Second, the *W*'s for the five-letter words are higher and lie between 0.68 and 0.76. Third, the correlations between the mean indices of the four- and the five-letter words vary between 0.75 and 0.83. And finally, there is no convincing evidence for the superiority of one index over the other.

Though the reliability coefficients are rather high, a closer look at the intercorrelation matrices revealed that they are deflated because of the weak correlation between the first series and the other of the same length. If the first series are dropped out, all *W* coefficients of the four-letter words increase with about 0.10 and range between 0.62 and 0.70 ($P<0.001$). The *W* coefficients for the five-letter words go up by about 0.12 and range between 0.79 and 0.88 ($P<0.0001$). However, the correlation between four- and five-letter words drops from about 0.78 to approx 0.50 ($P<0.05$), except for the *e* index which remains at 0.75. This drop is entirely due to the high intercorrelation between the first four-letter series and the five-letter series. The first five-letter series does not correlate significantly with the four-letter series, except with the first one. Intercorrelations between the different laterality indices remain roughly the same with or without the first series.

Two reasons can be invoked for the low correlation between the first and the subsequent series. First, it could be hypothesized that the low coherence is due to a lack of practice. In that case, the correlation between the first four-letter series and the remaining series will be higher for the subjects starting with a five-letter series than for the subjects starting with a four-letter series. Similarly, the correlation between the first five-letter series and the remaining will be higher for the subjects starting with a four-letter series than for the subjects starting with a five-letter series. Table 4, however, shows that this is not so. There is no interaction between the series started with and the correlation between "first series" and subsequent ones. Thus, the low correlation cannot be due to a lack of practice but probably is caused by unfamiliarity with the stimuli.

A last question we might ask about the reliability of the accuracy data is whether differences in recognition between different words influenced the magnitude of the correlation coefficients. As mentioned in the method section, there were three different types of four-letter words and two different types of five-letter words. Because the different word types are likely to involve different recognition rates (see below), it might be that the random assignment of words to the LVF and RVF sometimes favours one half-field and then the

Table 2. Spearman's rank correlation between the different laterality indices (average per subject) based on accuracy data. Four-letter words above the diagonal and five-letter words below the diagonal

	d	d_0	*POE*	*e*	*Phi*	*L*
d	—	0.91§	0.75‡	0.80‡	0.99§	0.86§
d_0	0.94§	—	0.53*	0.59*	0.90§	0.72†
POE	0.79‡	0.64†	—	1.00§	0.82‡	0.94§
e	0.84‡	0.71†	0.96§	—	0.86§	0.96§
Phi	0.98§	0.89§	0.85‡	0.91§	—	0.91§
L	0.94§	0.84‡	0.92§	0.96§	0.99§	—

*$n = 14$, $P < 0.05$.
†$P < 0.01$.
‡$P < 0.001$.
§$P < 0.0001$.

Table 3. Reliability of the laterality indices based on accuracy for the four-letter words and the five-letter words, and Spearman's rank correlation between the sum of the four-lettter words and the sum of the five-letter words. First series included

	Kendall's *W*		
	Four letters	Five letters	*Rho* Σ4Σ5
d	0.55†	0.76‡	0.76†
d_0	0.60†	0.75‡	0.79†
POE	0.54†	0.73‡	0.83†
e	0.55†	0.68†	0.79†
Phi	0.53*	0.75‡	0.75†
L	o.52*	0.70†	0.78†

*$P < 0.01$.
†$P < 0.001$.
‡$P < 0.0001$.

Table 4. The average *rho* between the *L*'s of the first and the four remaining series, and the average intercorrelation between the *L*'s of the four remaining series (*rho*) as a function of the series started with

	Started with four letters		Started with five letters	
	First series	Remaining	First series	Remaining
Four letters	0.25	0.25*	0.62	0.75†
Five letters	0.19	0.66	0.60	0.77†

*Low correlation caused by the unstable pattern of the first author who had written and tested the programs before taking part in the experiment.
†$P < 0.05$.

other and thus impairs test–retest stability. One way to assess this possibility is no longer to consider each word as having an equal weight (namely 1), but to take per subject the average between the recognition rate in the LVF and the recognition rate in the RVF as an estimate of the recognizability ($recogn_i$) of a word (w_i). By then dividing the sum of

$recogn_i \times w_i$ of all correct answers by the sum of $recogn_i \times w_i$ of all presented words, we get an estimate of accuracy corrected for overall recognizability. The words that were presented only in one half-field were left out of consideration. The thus corrected accuracy data did not yield better results than the uncorrected ones. The reliability coefficients for the different indices for the corrected data are quite similar to, and if anything, slightly lower than those for the raw data. So, the random assignment of words to the LVF and RVF does not have impact on the laterality indices derived from the accuracy data. Essentially the same picture is obtained when the first series are left out of consideration.

To measure the effect of the different word types, a $2 \times 3(2) \times 2 \times 5$ ANOVA with one between factor (Handedness) and three within factors (Word type, VHF, and Series) was calculated on the accuracy data of the four- and the five-letter words separately. Apart from the effects already known (main effect of VHF and Series), the ANOVA of the four-words yielded but two additional significant effects: a main effect of Word type (accuracy high frequency words = 0.75, low frequency words first letter = 0.59, low frequency words last letter = 0.61, $F(2, 24) = 39.317$, $P < 0.001$), and an interaction between Word type and Series (high frequency words = 0.69, 0.76, 0.73, 0.80, 0.78; low frequency words first letter = 0.52, 0.57, 0.58, 0.65, 0.65; low frequency words last letter = 0.46, 0.59, 0.62, 0.68, 0.67; $F(8, 96) = 2.975$, $P < 0.01$). The main effect of Word type is entirely due to the high frequency words being better recognized than the low frequency words. The interaction of Word type and Series is mainly caused by the steeper increase of accuracy over time for the low frequency words than for the high words.The ANOVA of the five-letter words yielded no significant effects apart from the main effect of VHF and Series, probably because the difference in frequency was not so large as for the four-letter words. The absence of a significant interaction between Word type and VHF, or between Series and VHF indicates that the RVF superiority was not affected by factors such as experience and repetition priming (Series), or word frequency and information distribution within a word (Word type). This was further confirmed by the absence of any significant effect in an ANOVA with Handedness, Word type, and Series as factors on the Lambda indices of the four- and the five-letter words. Especially the absence of a Word type × VHF interaction for the four-letter words is interesting, as it supports BRYDEN'S [5] statement that horizontal word display does not affect RVF superiority. If indeed the word-beginning is more visible in the RVF and the word-end in the LVF, we would expect infrequent words ending like a frequent word to be better recognized in the RVF, and infrequent words beginning like a frequent word to be better recognized in the LVF. This is not the case. Actually, the data tend to go in the opposite direction: low frequency words ending like a frequent word were better recognized in the LVF than low frequency words beginning like a frequent word (0.51 against 0.48), and worse in the RVF (0.68 against 0.73).

Latency data

As for the accuracy data, a $2 \times 2 \times 2 \times 5$ ANOVA with one between factor (Handedness) and three within factors (VHF, Wordlength, and Series) was done on the latency data of the correct trials. The factor Series again was random, the other factors fixed. A problem with the reaction time data was that the subjects sometimes made noise when the stimulus was being presented and thus triggered the timer before a voluntary reaction could have taken place. There were also a few occasions on which a subject waited too long before giving an answer. Though the answer was correct, these data distort the mean reaction time excessively. Therefore, all reaction times smaller than 200 msec and larger than 2500 msec were dropped from the analysis. Mean number of outliers per series ranged from 0 to 2.3. Neither ANOVA nor loglinear analysis are appropriate to analyse the differences in the number of outliers because the dependent variable is not continuous and the number of observations per cell is too small. Visual inspection, however, reveals that most outliers, not unexpectedly, were situated in the first series of the four- and the five-letter words (mean 1.41 against 0.51 for the other series).

Two main effects and one interaction of the ANOVA of reaction times were significant. For the main effects these were: VHF (LVF = 769, RVF = 693; $F(1, 12) = 34.93$, $P < 0.0001$), and Series (first presentation = 897, second = 731, third = 700, fourth = 687, fifth = 640; $F(4, 48) = 22.58$, $P < 0.0001$). The only interaction that reached significance was the interaction between Wordlength and VHF (four-LVF = 762, five-LVF = 776, four-RVF = 700, five-RVF = 686; $F(1, 12) = 4.76$, $P < 0.05$).

Because floor- and ceiling-effects are not so pervasive for reaction times as for accuracy data, only a few laterality indices have been proposed for reaction times. The major distinction which has been made is whether the variability of reaction times is inherent to cerebral asymmetry or not. In the latter case, variability can be discarded and a difference score (d) between the mean or the median time of LVF and RVF is appropriate. If, however, the variability is inherent to the nature of hemispheric differentiation, as LEVY [14] holds, it must be incorporated in the laterality index. A way to do this is to calculate the correlation between performance on each trial and sensory half-field (dichotomized as zero and one). This yields a point-biserial correlation (r_{pb}), related to the *phi* index of the accuracy data [14]:

$$r_{pb} = \frac{\overline{RT}_L - \overline{RT}_R}{s} (PQ)^{1/2}$$

in which RT_L and RT_R stand for mean naming latency in LVF and RVF, P and Q for the proportion of stimuli presented to the LVF and RVF, and s for the standard deviation of all stimuli presented.

The Spearman rank correlation between the average d (means) and r_{pb} per subject was 0.68 ($n = 14$, $P < 0.01$) and both indices yielded but one significant main effect in an ANOVA with one between factor (Handedness-fixed) and two within factors (Wordlength-fixed-, and Series-random). The significant main effect concerned the factor Wordlength (d: four-letter words = 62 msec, five-letter words = 89 msec, $F(1, 12) = 4.76$, $P < 0.05$; r_{pb}: four-letter words = 0.16, five-letter words = 0.25, $F(1, 9) = 5.52$, $P < 0.05$).

The reliability of the two indices is tabulated in the first part of Table 5. The W coefficients of concordance were significant only for the four-letter words, and were considerably lower than for the accuracy data. The correlation between the four- and the five-letter words is significant but higher for r_{pb} than for d. Similar, though slightly worse, results have been obtained for the d index based on the median reaction time (corrected for outliers). Reliability and correlation coefficients all improve if the first series are left out of consideration, as also can be seen in Table 5.

As for the accuracy data we might wonder whether the random assignment of words to the LVF and RVF could have a detrimental effect on the reliability scores. It is well known that naming latency for words varies according to factors such as word frequency and grapheme–phoneme regularities [20]. So, if there are considerable differences in naming latencies for the words used in this experiment, the alternation of words between LVF and RVF may considerably decrease test–retest stability. To test this possibility, an estimate of naming latency for each word and each subject was obtained by taking the mean of the naming latency for LVF presentation and RVF presentations. Words not recognized in one

Table 5. Reliability of the laterality indices based on naming latency for the four-letter words and the five-letter words, and Spearman's rank correlation between the sum of the four-letter words and the sum of the five-letter words. Data with and without first series

	Raw data					
	First series included			First series excluded		
	Kendall's *W*			Kendall's *W*		
	Four letters	Five letters	*Rho* Σ4Σ5	Four letters	Five letters	*Rho* Σ4Σ5
d	0.44†	0.27	0.47*	0.52*	0.43*	0.65†
r_{pb}	0.44†	0.27	0.64†	0.54†	0.49*	0.70†
	Data corrected for overall naming latency					
	First series included			First series excluded		
	Kendall's *W*			Kendall's *W*		
	Four letters	Five letters	*Rho* Σ4Σ5	Four letters	Five letters	*Rho* Σ4Σ5
d	0.44†	0.42†	0.29	0.67†	0.54*	0.69†
r_{pb}	0.50†	0.38*	0.58*	0.63†	0.69‡	0.41

*$P<0.05$.
†$P<0.01$.
‡$P<0.001$.

of the two visual fields were left out of consideration. Once the mean naming latency was determined d and r_{pb} were calculated on the difference between the reaction times and the mean latencies. The second part of Table 5 shows the results of the analysis.

Three things are noteworthy in comparing the two parts of Table 5. First, reliability scores for reaction times, just as for accuracy data, increase if the first series are excluded. Second, contrary to the accuracy data, test–retest stability is contaminated by the random assignment of stimulus words to LVF and RVF. And third, when irregularities of first presentations and naming latency differences are partialled out, the r_{pb} index appears to be superior to the d index. The correlation between d and r_{pb} remains the same after the corrections ($rho=0.65$, $n=14$, $P<0.01$) as before. As for the accuracy data, high reliability indices for the four- and the five-letter words go along with rather low intercorrelation.

Whereas the low correlation between the first and the subsequent series for the accuracy data was mainly due to a lack of familiarity with the stimuli, Table 6 indicates that for the r_{pb} indices, both factors seem to contribute partially. The point-biserial correlation tends to be higher for the "first series" presented secondly, but still remains lower than the intercorrelations further on.

Table 6. The average *rho* between the r_{pb}'s of the first and the four remaining series, and the average intercorrelation between the r_{pb}'s of the four remaining series (*rho*) as a function of the series started with

	Started with four letters		Started with five letters	
	First series	Remaining	First series	Remaining
Four letters	0.29	0.55	0.62	0.74*
Five letters	0.11	0.43	0.07	0.61

*$P<0.05$.

Word type did not have a significant influence as no effect involving it was significant in an ANOVA employing the variables Handedness, Word type, VHF, and Series. This was the case for the four-letter words as well as for the five-letter words. The absence of a significant VHF × Word type interaction for the four-letter words again supports BRYDEN'S [5] statement that, at least in our experiment and for the four-letter words, the horizontal display of words was not the main factor in the RVF superiority (high frequency words: LVF = 749 msec, RVF = 685 msec; low frequency words first letter: LVF = 775 msec, RVF = 706 msec; low frequency words last letter: LVF = 777 msec, RVF = 715 msec; $F(2, 24) = 0.052$, $P > 0.500$). The same was found for the r_{pb} index, as for both word lengths no effect reached significance in an ANOVA consisting of Handedness, Word type, and VHF.

The relation between accuracy data and reaction times

To see whether accuracy data and reaction times lead to the same results, Spearman's rank correlations between the lambda index of the accuracy data and the point-biserial correlation of the reaction times are tabulated in Table 7. Correlations were calculated on the average indices per subject for the four-letter words, the five-letter words, and their sum. The first series were not included. As could be expected, the correlation is highest for the sum of the four- and the five-letter words. However, the overall picture shows that the accuracy data of the four-letter words contribute very little to the correlation. This agrees with the rather low reliability scores of the four-letter words for accuracy.

Table 7. Spearman's rank correlation between lambda based on accuracy and r_{pb} based on naming latency. Data for the sum of the four-letter series, the five-letter series, and their combination. First series not included. Naming latencies corrected for overall naming latency

		Naming latency		
		Σ4	Σ5	Σ4Σ5
	Σ4	0.57*	0.57*	0.61*
Acc.	Σ5	0.55*	0.66†	0.76‡
	Σ4Σ5	0.69†	0.65†	0.81‡

*$P < 0.05$.
†$P < 0.01$.
‡$P < 0.001$.

DISCUSSION

Reliability studies usually provide a mass of data (not all of which have been presented in the Results section), certainly if the merit of different indices is to be investigated. This may look a little bit confusing at the beginning, but a closer look reveals that they all point to the same conclusions.

First, the study reveals that the first series of a visual half-field experiment involves strategies that, at least for the subject sample tested, seem to be unrelated to the strategies used in later series. This agrees with FENNELL *et al.* [8, 9] and thus care must be taken to give subjects sufficient practice before starting the real measurement if a stable pattern of results is to be obtained (e.g. for assessment). A discussion has arisen between MCKEEVER [15] and FENNELL *et al.* [8, 9] about whether the low correlation between first and later series is due to mere practice or to acquaintance with the stimuli. If the former is true, the correlation between the first four-letter series and the other should be lower for the subjects starting with a four-letter series in our experiment than for the subjects starting with a five-letter series. The same should be true for the correlation between first and subsequent five-letter series for the subjects starting with a five-letter series. As can be seen in Tables 4 and 6, only the reaction time data give some support to McKeever's practice hypothesis. The data are more in line with FENNELL *et al.*'s [8, 9] statement that it is unfamiliarity with the stimuli that causes the low intercorrelations. In practice, this means that the series of words are best presented more than once.

Accuracy and latency data lead to the same results and thus confirm BABKOFF and FAUST'S [1] position. Both can be used to assess cerebral dominance, either apart or in combination. However, some differences exist in our study. First, accuracy data are more reliable for five-letter words than for four-letter words (see Tables 3 and 7), while this is not the case for the reaction times (Tables 5 and 7). More data are needed to see whether this is a general finding or simply a result of our stimulus pool. Second, contrary to the accuracy data, reaction times are very susceptible to the random assignment of words with different naming latencies to the LVF and RVF. Therefore, either *a priori* estimates of mean naming latency for each word need to be collected, or the words must be presented in both VHFs in order to determine their mean naming latency (see Table 5). The latter will be more difficult if recognition rate is low.

We did not obtain much evidence for the superiority of one laterality index based on the accuracy data over the other (Table 3). Moreover, all laterality indices correlate very highly (Table 2) and lead to practically the same results (see ANOVAs on the indices). Therefore, the choice of index must be based on other criteria than reliability. Personally, we feel most positively about SPROTT and BRYDEN'S [4, 24] lambda-index. It correlates highest with all the other indices (Table 2), is assumed to be unsusceptible to floor- and ceiling-effects [4, 24], and allows an analysis of each subject separately [6].

For naming latency, a superiority of the point-biserial correlation over the mere difference index is obtained when the first series are dropped out and the differences in naming latency between the words are corrected for (Table 5). This supports LEVY'S [14] dynamic model of functional hemispheric differentiation, which holds that at least part of the variation in naming latencies is due to factors related to hemispheric differentiation, such as general activation, attentional focusing, and degree of distraction. An additional advantage of the point-biserial index over the difference index is that it allows individual analysis just like the lambda index [6].

Five-letter words, in agreement with the existing literature [27], yield larger asymmetry indices than four-letter words. The results are unaffected by word frequency and information distribution across the word, at least in our experiment where stimulus presentation was not blocked. The only effect we obtain is that the recognition rate drops as word frequency becomes very low (see four-letter words). Things may be different for other variables, such as word imageability [16, 27].

CONCLUSION: RECOMMENDATIONS FOR PRACTICAL USE

Though it is always a bit equivocal to generalize data from one study, some recommendations and precautions may be formulated on the base of the results just presented. Hopefully, future experiments will make things more firm.

There is evidence that different processes and strategies are involved in the first and subsequent stimulus series. This implies that more than one series must be presented if a stable asymmetry level is needed (e.g. for assessment or in case the experimenter wants to correlate cerebral asymmetry with another variable). It also seems that the difference between the first and subsequent stimulus series is not due to mere practice but to acquaintance with the stimuli, so that the same or very similar stimulus series must be used. Things may be different for group comparisons, as no ANOVA yielded a VHF × Series interaction.

Accuracy and naming latency lead to the same results. They can be used together and provide converging evidence, but this is not a necessity. The study above, however, provides some evidence that if accuracy is the major dependent variable, greater reliability is obtained with five-letter words than with four-letter words, though further research may be needed to make sure that the finding is not caused by the stimulus sample used. Word frequency does not seem to have a great influence, so that the experimenter is rather free to chose the stimuli as for this variable. The only thing to be kept in mind is that the accuracy drops as the stimuli become too infrequent. To ensure that the horizontal display does not influence the RVF superiority, a manipulation like the one used in this study can be incorporated. An alternative would be to use stimuli that can easily be guessed from the beginning or ending letters, and to check whether there is a difference between the LVF and RVF. Other variables have not been checked in this study.

Laterality indices based on naming latency are influenced by the choice of stimuli in the LVF and RVF. This can be bypassed by prior experimentation that ensures the equivalence of the stimulus samples presented to the LVF and RVF, or by counterbalancing. The latter option will increase the variability and asks for a double presentation of a series (left and right) if individual assessment is needed. The need to control the stimuli that are presented to a VHF, makes pure randomization (with replacement) as used in this article less interesting. A better way is first to categorize the stimuli according to the VHF in which they are to appear, and then to permute the whole series in a random fashion (i.e. randomization without replacement, see [17]). This will greatly simplify calculations.

The point-biserial correlation coefficient was found to be a more reliable laterality index of naming latency than the mere difference between LVF and RVF. No such superiority was obtained for the indices based on accuracy data. However, because of its elegant statistical properties, we would prefer the lambda index.

Finally, the whole points to the need to present a large number of stimuli to each subject. The difference between the first and the subsequent series, the influence of word variables on the indices based on naming latency, and the lack of complete reliability, all urge each clinician and experimental psychologist to invest a fair amount of energy in a VHF study (i.e. certainly more than one session of 100 stimuli per subject). Only then will the results be embedded in sufficiently small confidence regions (see also [6]) and stable enough to be an indication of the lateralization of the subject.

REFERENCES

1. BABKOFF, H. and FAUST, M. Lexical decision and visual hemifield: an examination of the RT-Accuracy relationship. *Neuropsychologia* **26,** 711–725, 1988.
2. BRADSHAW, J. L. and NETTLETON, N. C. *Human Cerebral Asymmetry*. Prentice-Hall, Inc., Englewood Cliffs, 1983.
3. BRADSHAW, J. L., NETTLETON, N. C., PIERSON, J. M., WILSON, L. E. and NATHAN, G. Coordinates of extracorporeal space. In *Neurophysiological and Neuropsychological Aspects of Spatial Neglect*, M. JEANNEROD (Editor), North Holland, Amsterdam, 1987.
4. BRYDEN, M. P. *Laterality: Functional Asymmetry in the Intact Brain*. Academic Press, New York, 1982.
5. BRYDEN, M. P. On the possible dangers of using horizontal word displays in visual field studies. *Brain Cognit.* **5,** 362–368, 1986.
6. BRYSBAERT, M. and D'YDEWALLE, G. Individual analysis of laterality data. *Neuropsychologia*, in press.

7. DAGENBACH, D. Subject variable effects in correlations between auditory and visual language processing asymmetries. *Brain Lang.* **28**, 169–177, 1986.
8. FENNELL, E. B., BOWERS, D. and SATZ, P. Within-modal and cross-modal reliabilities of two laterality tests. *Brain Lang.* **4**, 63–69, 1977.
9. FENNELL, E. B., BOWERS, D. and SATZ, P. Within-modal and cross-modal reliabilities of two laterality tests among left-handers. *Percept. Mot. Skills* **45**, 451–456, 1977.
10. GRANT, D. W. Sex, reading ability, and visual half-field test–retest reliability. *Percept. Mot. Skills* **54**, 49–50, 1982.
11. HERON, W. Perception as a function of retinal locus and attention. *Am. J. Psychol.* **70**, 38–48, 1952.
12. HINES, D., FENNELL, E. B., BOWERS, D. and SATZ, P. Left-handers show greater test–retest variability in auditory and visual asymmetry. *Brain Lang.* **10**, 208–211, 1980.
13. KIRSNER, K. and SCHWARTZ, S. Words and hemifields: do the hemispheres enjoy equal opportunity? *Brain Cognit.* **5**, 354–361, 1986.
14. LEVY, J. Individual differences in cerebral hemisphere asymmetry: theoretical issues and experimental considerations. In *Cerebral Hemisphere Asymmetry: Method, Theory, and Application*, J. B. HELLIGE (Editor), pp. 465–497. Praeger, New York, 1983.
15. MCKEEVER, W. F. Tachistoscopic methods in neuropsychology. In *Experimental Techniques in Human Neuropsychology*, H. J. HANNAY (Editor), pp. 167–211. Oxford University Press, New York, 1986.
16. MCMULLEN, P. A. and BRYDEN, M. P. The effects of word imageability and frequency on hemispheric asymmetry in lexical decisions. *Brain Lang.* **31**, 11–25, 1987.
17. NILSSON, T. H. Randomization without replacement using replacement without losing your place. *Behav. Res. Meth. Instr.* **10**, 419, 1978.
18. PORAC, C. and COREN, S. *Lateral Preferences and Human Behavior*. Springer-Verlag, New York, 1981.
19. ROSZKOWSKI, M. J. and SNELBECKER, G. E. Validity and temporal stability of the chimeric face technique for studying hemispheric processing: data from 6- through 14-year-old children. *J. Behav. Ass.* **4**, 209–221, 1982.
20. SCHEERER, E. Visual word recognition in German. In *Language Perception and Production*, A. ALLPORT, D. G. MACKAY, W. PRINZ and E. SCHEERER (Editors), pp. 227–244. Academic Press, London, 1987.
21. SCHWARTZ, S., MONTAGNER, S. and KIRSNER, K. Are there different methods of lexical access for words represented in the left and right visual fields? *Brain Lang.* **31**, 301–307, 1987.
22. SIEGEL, S. *Nonparametric Statistics for the Behavioural Sciences*. McGraw-Hill, New York, 1956.
23. SPRINGER, S. P. Dichotic listening. In *Experimental Techniques in Human Neuropsychology*, H. J. HANNAY (Editor), pp. 138–166. Oxford University Press, New York, 1986.
24. SPROTT, D. A. and BRYDEN, M. P. Measurement of laterality effects. In *Cerebral Hemisphere Asymmetry: Method, Theory, and Application*, J. B. HELLIGE (Editor), pp. 445–464. Praeger, New York, 1983.
25. UIT DEN BOOGAERT, P. C. (Editor) *Woordfrequenties in geschreven en gesproken Nederlands*. Oosthoek, Scheltema & Holkema, Utrecht, 1975.
26. WOLACH, A. H. and MCHALE, M. A. *F* ratios and quasi *F* ratios for fixed, mixed, and random model ANOVAs. *Behav. Res. Meth. Instr. Comp.* **19**, 409–412, 1987.
27. YOUNG, A. W. and ELLIS, A. W. Different methods of lexical access for words presented in the left and right visual hemifield. *Brain Lang.* **24**, 326–358, 1985.
28. YOUNG, A. W. and ELLIS, A. W. More things in heaven and earth than are dreamt of in the initial letter acuity hypothesis. *Brain Lang.* **31**, 364–371, 1987.