0196-0202/82/0306-0340$02.00/0
EAR AND HEARING

Vol. 3, No. 6
Printed in U.S.A.

LETTERS TO THE EDITOR

Comment on "A Quantitative Examination of the Sources of Speech Discrimination Test Score Variability"[1]

Aaron Thornton, and Michael J. M. Raffin

Harvard Medical School, AER Laboratory, MEEI, Boston, Massachusetts [A. T.], and University of Montana, Department of Communication Sciences and Disorders, Missoula, Montana [M. J. M. R.]

We wish to compliment Dr. Dillon on his well organized and clearly presented discussion on speech test variability. Although he has been careful in subjecting our data[2] to two different binomial models, we would like to correct one error and emphasize that the binomial model presented in our paper was not discussed by Dillon.

An inference was made on page 53 that our experiment was of the type "where each subject has heard a list two or three times, with differences in scores between test and retest computed individually for each subject."[1] On the next page Dillon states that "the goodness of fit of the simple binomial model must be due in part to there being variability in addition to that expected on binomial grounds." Our test-retest data were not of this type, and our simple binomial model is significantly different from the one described by Dillon. Although the formula for computing variance is the same, Dillon and we are modeling different processes.

Our study was specifically concerned with intrasubject variability across test forms, which is the condition most frequently encountered by clinicians using open response tests. Conversely, Dillon dispensed with the influence of list differences early in the paper and focused on the problem of modeling variance across repetitions of the same test forms. The "simple" binomial model underestimated this source of variance. However, our subjects heard each word only once, and variability across lists, rather than variability between presentations of the same lists, was evaluated. When different lists are used, a simple binomial model based on sampling theory may be applied which neither presumes equally difficult items nor requires knowledge of individual word difficulties. Consequently, the goodness-of-fit for our data should be expected. The additional variability that Dillon seeks can be found in the within-subject variance across test forms which is not part of either of his binomial models.

Although test forms can be easily adjusted to produce equivalent scores averaged across subjects, it is very difficult to produce forms which will be truly equivalent for each of a large group of subjects, particularly if one includes a large range of ages, hearing losses, nationalities, listening conditions, etc. If complete equivalence were achieved for each subject, then the factors determining test-retest variance would be the same as those in which the same words are given repeatedly, and our sampling model would overestimate variance. Currently, none of our open set speech discrimination tests appear to be very close to having truly equivalent forms.

References

1. Dillon, H. 1982. A quantitative examination of the sources of speech discrimination test score variability. Ear Hear. **3**, 51–58.
2. Thornton, A. R., and M. J. M. Raffin. 1978. Speech discrimination scores modeled as a binomial variable. J. Speech Hear. Res. **21**, 507–518.

The Sources of Speech Discrimination Test Score Variability: a Reply to Thornton and Raffin

Harvey Dillon

National Acoustic Laboratories, Sydney, Australia

I would like to thank Drs. Thornton and Raffin for pointing out an error in my paper.[2] It is certainly true that the estimates of test-retest reliability obtained in both their research[4] and that of Hagerman[3] include a component due to interlist differences. I am also in complete agreement with their comment that it is difficult to produce open response-set test forms which are

truly equivalent for each subject. (This, of course, was one of the major motivations for the development of closed response-set tasks.)

However, I cannot accept their claim that when test and retest are based on different lists, the simple binomial model used by them may always be applied, irrespective of the mixture of item difficulties within each list. To illustrate this, let us perform the following armchair experiment. Suppose we have two lists, of 50 words each. Twenty-five of the words on each list are exceptionally easy, while 25 are exceptionally hard. Each subject undergoing the test thus scores 50% on each list, and we predict from the simple binomial model and formula that the standard deviation of the distribution from which each score is obtained should be 7%. This is equivalent to a standard deviation for the test-retest differences of 10%. By contrast, the observed test-retest difference will always be exactly zero for these lists, as correctly predicted by the subnormal binomial distribution[2] which takes the item difficulty mixture into account.

In the simple binomial model developed from sampling theory by Thornton and Raffin,[4] this particular combination of test word difficulties is of no consequence, as the next time this or some other subject is tested, the test form will be comprised of a different random combination of easy and hard words. In the real clinic, however, only a finite number of fixed lists is usually available, and so the effect of a mixture of item difficulties on test variability remains.

Clearly, the example portrayed above represents an extreme case, but it is one which would be possible to at least approximate by various means and is therefore one which provides a useful test of any theoretical model. Less extreme mixtures of item difficulty will be accompanied by less extreme deviations from the simple binomial predictions. The presence, in existing fixed lists, of words which are "easy" and "hard" on the average, as well as for each individual subject, is something most clinicians become aware of, and is also well documented.[1]

In view of the above discussion, I would like to reiterate and slightly modify my previous conclusions pertaining to intrasubject test-retest reliability.

1. The simple binomial model can overestimate variability due to a mixture of item difficulties in the test.
2. Either binomial model can underestimate variability due to excessive fluctuations in performance by some subjects, and due to interlist differences if applicable.
3. Because of the opposing effects of the deviations from the model, the simple binomial formula provides a good estimate of the total average intrasubject variability.

References

1. Campbell, R. A. 1965. Discrimination test word difficulty. J. Speech Hear. Res. **8**, 13–22.
2. Dillon, H. 1982. A quantitative examination of the sources of speech discrimination test score variability. Ear Hear. **3**, 51–58.
3. Hagerman, B. 1976. Reliability in the determination of speech discrimination. Scand. Audiol. **5**, 219–228.
4. Thornton, A. R., and M. J. M. Raffin. 1978. Speech discrimination scores modeled as a binomial variable. J. Speech Hear. Res. **21**, 507–518.