Psychological aspects of Snellen VA scores

C. French*, S. Laverty, R. Pyke, G. Ridgewell, and P. Vitkovitch

The following report is based on experimental projects carried out by four of the authors in the final year of an honours degree course in ophthalmic optics as part of a psychology option. Chris French is a lecturer in the department of optics.

During an eye examination the patient is subjected to a battery of tests and questions, the reliability and validity of which may be profoundly affected by psychological factors such as attention, learning, memory, suggestibility and intelligence. The object of the present study was to investigate the influence of practice on a conventional measure of'visual acuity using a new approach provided by computer-drawn Snellen charts. The experiments were carried out as psychology projects by four of the authors during their final year in an honours degree course in ophthalmic optics.

One problem with Snellcn charts is that they may contain a limited number of lines with the result that the patient may read the same lines of the chart again and again. It is impossible that practice alone will lead to an apparent improvement in visual acuity. It is also possible that a subject will memorise letter strings, making it possible for him to recite the lines without reading every letter. We did not want to confound these two effects so we constructed a new test made up of special charts which eliminated the possibility of memory being used and guaranteed that our task was one of letter identification alone.

Each test consisted of around 20 charts each made up of six lines containing six letters. The ratio of the size of the letters on the top line to those on the bottom was six to one. Each letter was chosen at random by a computer programme which then instructed a microfilm plotter to draw it in the appropriate size and position directly on 35mm film. A set of charts was complete when every letter of the alphabet had appeared at least once in each of the six lines. In this way we had a means of obtaining VA — type scores for all the letters from A to Z

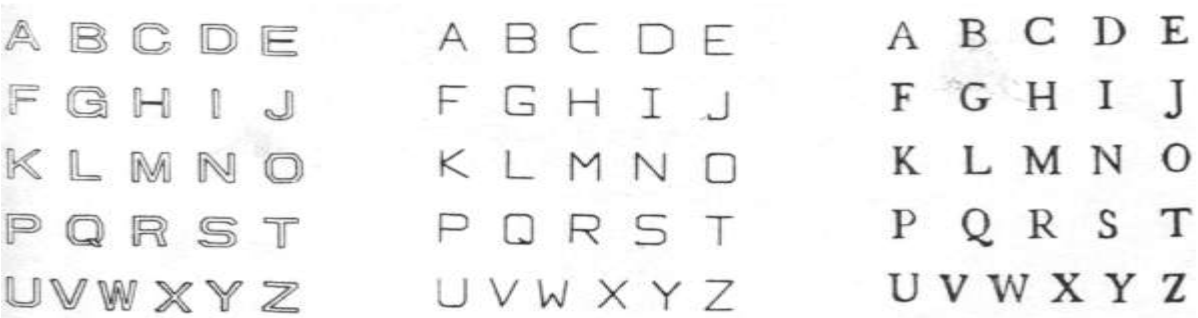
Several type faces or founts were used which were readily available "off the computer-peg" [1-2]. We used the three illustrated in Fig 1. They are, from left to right, a plain outline fount (fount *1):* a plain single line fount (fount 2); and a serif fount with each limb of the letter made up of several strokes (fount 3).

The charts were projected by a slide projector onto a screen placed 3 metres from where the subjects sat. The instructions for the subjects were that they had to identify each letter on every chart from the top to the bottom.

If the subjects were not sure what a letter was, they had to guess. By using this "forced-choice" method it was possible to avoid any change in their success rate being attributable to merely an increase in their confidence. The score of a subject was the angular height of the smallest example of a letter which the subject could read. Only a letter's first occurrence on a particular line in a set of charts counted for this purpose. Viewing conditions were constant within each particular experiment.

^Experimental psychologist. **UMIST**

Figure 1. The three founts used in the experiments reported



Experiment 1—Transfer of practice effects from one fount to another.

The first experiment sought to compare the legibility of founts 1 and 2 and see whether practice on one fount would lead to improved acuity scores on the other. Transfer of training* is a familiar phenomenon and it would appear not unreasonable to expect practice insofar as it may lead to an improvement in one perceptual skill to also lead to an improvement in a similar one.

The subjects were 24 secondary schoolchildren who were tested using both founts. Our main hypothesis was not supported: practice did not lead to any statistically significant improvement. Ignoring temporal changes, a comparison of the scores for the two founts revealed that on average there was no difference in their legibilities, although a more detailed statistical analysis showed that for some letters (A, I, K, M, V, W and X) fount I was more legible while for others (C, F, G and S) the reverse was true. The angular size of the smallest letters successfully read by the average subject varied from 3.0 minutes of arc for the letters I and L to 7.4 minutes for the letter G (both in fount I).

Experiment 2—Experimenter effects, intermodal effects and mastication

It has been suggested that chewing gum can be of assistance in helping people do things a little better. This may well be true with perceptual tasks such as visual acuity for two main reasons: (i) chewing might improve concentration and thereby reduce errors psychologically; (ii) experimental work (notably by Russians[3]) has shown that additional sensory stimulation may heighten arousal and thereby improve perceptual performance in another modality.

The second experiment was designed to test this out and see whether sensory feedback from mastication could improve letter acuity scores. Two main groups were established. The first was a control group whose performance was simply measured on three occasions. The second was an experimental group. These, too, were tested on three occasions but on the second and third times they were given gum to chew.

A major source of error in psychology experiments can be labelled 'experimenter effects'. The experimenter's own beliefs or hypotheses have frequently been shown to effect the behaviour of his subjects and thereby his results.[4] Because of this it is normal practice not to tell subjects the purpose of any experiment they are participating in.

Even if the purpose of an experiment is not spelled out subjects are likely to form their own hypotheses as to its purpose and this conjecture may well influence the results. Thus, in the current context it would not have been long before the gum-chewing subjects came to their own conclusions as to the purpose of the study.

To control for this possibility the experimental group was divided in two. Half the subjects were told that gum improves vision, while half were told it interferes with it. We thus had an experimental design which was capable of not only establishing the effects of gum chewing but also whether or not these results could be affected by suggestion.

The results of the experiment were quite clear. There was no statistically significant difference between the three groups of subjects and no significant improvement in VA scores over the three sessions. This, then, is the second experiment to fail to reveal any performance improvement with practice. On the question of chewing gum it would appear that an optician need only fear it when it was out of a child's mouth!

Experiment 3—The influence of practice with and without feedback

Previous research has shown that practice *can* lead to an improvement in acuity performance.[5] The nearest experiment to our own is perhaps one which used an E chart with illumination which could be progressively lowered. The threshold (determined by the minimum illumination at which the subject could detect the orientation of the E) was found to be lowered by repetition or practice. However, sensitivity was raised even more when feed-back in the form of knowledge of results was also given.

In the third experiment, unlike the other two, the subjects used were students. Four subjects received knowledge of results feedback - - whenever they made an incorrect guess the experimenter pointed this out to them and at the same time told them what the correct response was. Four other subjects performed without such feedback.

Testing was carried out on five separate occasions on different days. It was thought that the long period of practice (five days as against two and three with the previous experiments) and the inclusion of feedback with one group would surely lead to an improvement in measured acuity.

Unfortunately a mistake † was made in setting up the experiment and the subjects were seated too close to the projector screen. As a result of this some subjects were successfully identifying almost every letter projected. In other words there was virtually no room for improvement what is sometimes called a "ceiling effect'. Paradoxically,

†The view has been expressed that students should not be entirely prevented from making mistakes in project work as these can frequently have educational value and one can also never completely rule out serendipity from playing a useful role!

despite this, improvement *did* occur and not just for the feedback group. The practice alone croup also improved! In fact, from a statistical point of view there was no difference between the two groups.

Experiment 4—Reading speed and the effect of practice

The notion behind the final experiment was that if practice did lead to improvements in performance, then it might be reasonable to relate this to the degree of practice which the subjects had already experienced in reading letters. In other words it might be that people who read a lot and were particularly practised in reading would have less room for any improvement from further practice, while those who read only a little would have more scope. Speculatively it could be argued that those who read a lot would be identified as fast readers whereas those who read a little would be slow readers. Clearly, even if our hypothesis is true, common-sense tells us that the magnitude of am such effect is unlikely to be great.

It is a well-known statistical fact that in order to establish subtle differences as being 'statistically significant' it is essential to have large groups of subjects in order that we may have reliable measurements of group means and such like. Unfortunately, this was not possible here for practical reasons and the archetypal psychology experiment was attempted with only four subjects in each group! Group one contained the 'fast' readers (more than 275 words per minute) and group two the 'slow' (less than 275 words per minute). It might be argued that even our labels for the two groups were misnomers as their reading speeds did not differ very markedly.

Conditions, then, were far from ideal for demonstrating any acuity differences between the groups so it was very surprising when the results revealed just that! The subjects were tested on three separate occasions in groups and with their spectacles on. Improvements with practice were shown for the letters B, D, G, M, N, O, Q, R, S, W and Z for both groups at levels of statistical significance from $p<0.05$ to $p<0.001$. At the same time the difference between the groups revealed itself in improvements for the slow readers over sessions for the letters C, H, K and Y <$p<0.05$ and $p<0.01$), while there was no such improvements for the fast readers.

Our reaction is to regard the empirical difference between the two groups with caution despite it being in the direction of our hypothesis. It is possible that the experimenter had unconsciously influenced the subjects and this was a late manifestation of the experimenter effect[5] we had been searching for earlier. The practice effect itself, however, cannot be dismissed in such a fashion—it is far too substantial. We had a reliable improvement over three sessions for at least eleven letters although a similar experiment (experiment 2) also using fount 3 and more subjects had negative findings. The average size of the smallest letters identified decreased from 3.6 to 2.7 minutes of arc.

Resolving differences between the experimental results

In considering how and why practice improves performance we have in the above results a fundamental problem. Experiments 1 and 2 did not show any such effect, while experiments 3 and 4 did and, indeed, experiment 3 contained conditions which it might be expected were least favourable towards showing such changes. As suggested before it is easy to 'blame' the person carrying out the experiment.

Each experiment was carried out by a different experimenter so obviously we have different results. But this is too facile. They each used the same computer technique and similar statistical methods and if we are merely to ascribe the different results to factors such as 'subtle* experimenter effects we might as well give up all research.

The saving explanation perhaps lies in the subjects. It so happens that the first two negative studies used schoolchildren in their early teens while the latter studies which showed positive improvements used undergraduates. The differences between the two classes of subjects are many. Apart from age it is likely that the students were more intelligent and this appears to us to be a possible source of the discrepant results, although a difference in the relationship between the experimenter and the subjects could equally be a candidate. The group being tested by its peer might well become more involved and try harder.

Psychologists have been criticised in the past (and perhaps still are) for apparently experimenting mainly on rats and pigeons, and then attempting to generalise their results to other animals and humans. Another criticism might equally be made that when psychologists (and other scientists) actually do get around to using human subjects they tend to concentrate exclusively on students—the human equivalent of the laboratory rat! There are, of course, good reasons for these strategies: a lot is known about the biology of the rat, it is a docile animal, can be kept under controlled conditions and is in plentiful supply. Similarly for most students: their academic backgrounds are well known, they tend to be tolerant to mild abuse, they live in standard student communities, and are also readily available and cheap to run. Both are a "convenience' animal as far as the experimenter is concerned. On the other hand perhaps neither is particularly representative either of animals or humans. It is difficult enough in psychology to establish empirical relationships between variables which hold for a single human being, let alone for a group of heterogeneous human beings of different backgrounds, intelligence, personality, sex, size and shape. What these four experiments show is how important such factors may be in something so apparently simple as measuring the effects of practice on perceptual performances.

Conclusions

Previous research has suggested that the checkerboard pattern provides the best measure of what has been termed 'retinal resolution'[6]. The Snellen chart, however, has been found to be a particularly good measure of another form of visual acuity—"form perception'[6]. In the four experiments reported we have been looking mainly at one source of error in the measurement of letter acuity — that due to practice. We have demonstrated that it can occur but that the prediction of how and when and how much is not easy.

In clinical practice there are also other sources of error which are probably more important. The patient may well remember the sequence of letters or a particular line may contain an easier than average set of letters. One would therefore expect measurements made in ophthalmic practice to be more unreliable.

Fig 2 shows the computer drawn letters placed in order of increasing difficulty alongside the rankings found by Coates and Woodruff for non-serif letters (discussed in detail by Bennett[7]). The only way for each line of a chart to be equally difficult is to include every letter in question in every line as was done in the present study. In practice this becomes practical only if a small set of letters is used in the chart constructions.

The time to present it is then manageable but the resulting test is less representative of letter acuity and more unreliable[6]. Still, despite the criticisms made of the conventional chart, it does appear to carry out its task quite well. Its test-retest reliability (repeatability of the measurements expressed as a correlation coefficient) is of the order of 0.88 in certain circumstances[6].

For the students who carried out these experiments their main value would appear to lie not in the individual results but in the careful experimental scrutiny of measurements of a type they will be making presumably for the whole of their professional career until they are (like the psychologists) replaced by a computer. A similar view was expressed by Mercer[8] who also considered that the most valuable way that psychology could be presented to ophthalmic optics students was via the medium of the experiment.

There are many other classes of measurements and observations that an optician will be making and each is subject to similar sources of what might be termed 'psychological' errors. It does not follow that because these errors are statistically significant they necessarily will be vital or even large. It would seem to be useful that an optician should regard the absoluteness of all his measurements with reservation, and if education can foster this critical attitude it will have done well.

Figure 2

Source of fount	fount	letters in order of decreasing recognisability.
expt. 1	plain	L U JO A I V C H X D E N P F T S Z Y B R G K W M Q
expt. 1	Helvetica	I L A U J V X O N P T W K Z E D Y R M F H B C S Q G
expt. 2	serif	L A O I Q W J T D C N V M X U R F P S K E Z B H Y G
expt. 4	serif	A J Q W L O C I N U T P D M X V R S F K E Y H B Z G
Coates	non-serif	A N V D P Z R Y X Q M B
Coates	non-serif	L T A E Z F H P N D V R
Woodruff	non-serif	L T U P Z E N H O D V C B Y G

Letters for the three founts used in the experiments arranged in order of difficulty along with Coates' and Woodruff's rankings.

References

1 M. V. Mathews, C. Lochbaum and J. A. Moss, Three founts of computer drawn letters. Communications of the ACM 1967, *10*, 627-630.

* F. R. A. Hopgood and D. C. Toll, *GROATS' Manual*. Chilton, Berkshire Atlas Computer Laboratory.

3 I. London, Research on sensory interaction in the Soviet Union, *Psychological Bulletin, 1954,* 5/f 531-568.

4 R. Rosenthal, Experimenter effects in behavioral research, New York: Appleton-Century-Crofts, 1966.

5 E. J. Gibson, Principles of perceptual learning and development. New York: Appleton-Century-Crofts 1969, pp 163-191.

6 U.S. Department of the Army, TAGO, Personnel Research Branch; Studies in visual acuity.*PRS Report* No. 742, 1948.

7 A. G. Bennett, Ophthalmic test types, *British Journal of Physio logical Optics*, 1965, 22, 238-271.

8 J. C. Mercer. A course in psychology for ophthalmic optics students. *The Ophthalmic Optician,* June 24, 1972. 584-587.