capstract:2610.00003v1 [med.CL] 27 Oct 2026

Original Investigation

Assignment of a disputed charge and near reading performance after cataract surgery: a randomized crossover trial

Livia Romeri[1,†], Serafim Danciu[2], Hanae Tersel[1,3], Olav Mikkelsen Vist[3]

[1] Department of Visual Function, Valderi Eye Institute, Parma, Italy
[2] Outcomes Unit, Trestin Center for Clinical Measurement, Cluj-Napoca, Romania
[3] Section for Clinical Optics, Vestdal Research Infirmary, Aarhus, Denmark

Correspondence: Livia Romeri, l.romeri@valderi.example

Abstract

Objective: To estimate the effect of assigning a disputed restaurant charge to the reader on the accuracy of near reading after cataract surgery.

Design: Two-period, two-center randomized crossover trial. Adults attending a scheduled postoperative visit read 35 amounts from itemized bills under each of two settlement instructions. A verified charge was returned either to the reader’s supplementary allowance or to the restaurant account. Print, viewing distance and time limits were held constant. The prespecified outcome was the number of amounts transcribed exactly.

Results: All 120 randomized participants completed both periods. Mean scores were 27.40 of 35 for the personal account and 23.68 for the restaurant account. The mean paired difference was 3.72 amounts (95% CI 3.14 to 4.30). At the largest print size the difference was two correct readings in 600 presentations. At the smallest size it was 101 in 600. Under concealed settlement instructions, 399 of 600 personal and 395 of 600 restaurant amounts were correct.

Conclusions: Who received the refund affected the reading of the charge. A change in performance equivalent to about three quarters of a chart line occurred within a single postoperative visit and depended on information supplied before the chart appeared.

Keywords: near vision; reading assessment; cataract surgery; crossover trial; task engagement

Introduction

In the postoperative clinic a patient sometimes brings a receipt to demonstrate the size of print they can now read. The receipt is folded along the item the patient wants the surgeon to see. Some patients can read the disputed entry from further away than the neighboring entries, including entries struck from the same roll of paper. The distinction matters when a sample of everyday print is used as evidence of the functional benefit of surgery. It also matters whether the patient has already established what the amount ought to have been.

Clinical reading charts control variation in spacing, contrast and number of characters. A sequence of geometrically spaced sizes makes the difference between two adjacent lines interpretable across the chart[1–3]. The text itself introduces other differences. A reader may complete a familiar word on the basis of two letters. Reading speed can change well above the smallest size at which a word can be identified[4–6]. Receipts contain few complete sentences and many repeated shapes, but their entries carry consequences which vary within a single sheet of paper. A second bottle of mineral water is of particular interest if it has been charged to the reader. The optometric properties of the second bottle are otherwise those of the first.

The account to which an amount is posted has generally been left unspecified in studies of everyday print. Work with supermarket labels found that participants could often reconstruct the price from their usual purchases[7]. In another series, patients who brought their own bills tended to bring bills they had successfully disputed[8]. Neither design could separate the reading of a number from recollection of it. We used

unpublished prices and transferred the right to a refund while leaving the printed bill in place. Our question concerned exact reading within a fixed interval. Neither optical resolution nor the clinical result of the operation was expected to change between accounts.

Methods

Patients and allocation

We enrolled patients at scheduled visits in Parma and Aarhus between March and September 2025. Eligibility required adulthood, cataract extraction in at least one eye 4 to 12 weeks earlier, and permission from the examining clinician to remain after the visit for a binocular reading assessment. We tested people with the near correction they had brought that day. An examiner asked each patient to read two large practice amounts and to indicate who would receive the difference if the amount was successfully disputed. An interpreter could explain the account rules. Patients had to read the digits themselves. We did not ask people who were staying for an urgent assessment to postpone it. A second visit by the same patient did not qualify as a second enrollment.

An independent statistician prepared the order of the two accounts within each center. Sets of four allocations balanced first account and first book of bills. A center received the next numbered envelope after the examiner had entered the results of the practice items. The room examiner could see the cards with the instructions. The scorer received only the response recordings and item identifiers, with period and account stripped from the file. Staff outside the room settled the accounts after the last observation. There was no interim examination of the scores. The crossover design and reporting plan were

recorded in the Valderi Clinical Methods Register (VCM-25-014) before the first study visit[9,10].

The disputed charge

Each receipt showed a department heading, three food entries and a small boxed amount on the last line. A duplicate cover had been added to a table of one. The boxed amount was the excess to be refunded. The question in both periods was what amount appeared inside the box. The number was new on each presentation. Neither the other prices nor the printed total allowed the reader to calculate it. The total was folded under in both books, a point settled in development when one reader subtracted two legible items to recover an illegible third. We retained the restaurant layout because it located the disputed entry without an arrow and supplied a reason for checking the paper[11].

The expense office opened a personal research allowance and a restaurant account. It reserved the same maximum supplementary payment in each. On a personal bill a correct reading released the disputed amount to the participant. On a restaurant bill it released the same amount to the restaurant’s study account. This account was held centrally and bought refreshments for subsequent clinic sessions. There was no individual restaurant owner at the assessment. The participant received a fixed attendance payment in all cases. No entry could reduce this payment. Each account was settled on a single bill drawn at random from its book at the end of the visit. A person who could read the first bill had the same incentive to read the last. We explained the draw with a large-print example and withheld feedback until both books had been returned.

The face of the bill bore no indication of its owner. Before each period, the examiner gave the participant the corresponding account card and asked who would receive a refund from that book. The examiner kept the card upright beside the book, outside the field of small print. The personal card carried a plain circle and the restaurant card a square. The personal pronoun appeared on the card in the participant’s preferred language. Both sites used the decimal comma. An amount was always two digits followed by a comma and two further digits. Units were study credits, converted to a local payment under a rate supplied in advance. The first and last digit varied independently of account. Values ending in 00, 50 or 99 were excluded because of their frequency on actual menus[12].

Presentation and scoring

Two books carried different amounts drawn from the same lists of digits. We matched leading digit, number of enclosed spaces in the printed numerals, and position of a repeated digit between books at each size. Each participant saw both books once. The optical shop printed the books in the same run on matte stock. A reference patch in the margin allowed staff to reject a print run for poor black density before binding it. The larger print used the same outlines as the smaller print. We kept the line position and the white margin in proportion to digit height. A faint rule that disappeared at the smaller sizes in the prototype had acted as a clue to size and was omitted[13].

The 35 items in a book comprised five at each of seven sizes, in steps of 0.1 on the common logarithm of digit height. The reference height at zero subtended five minutes of arc at the test distance. The largest was 0.7 and the smallest 0.1 on this scale. A numeral in this font does not have five equal critical details. We therefore report log digit

size and reserve logMAR for the clinical letter score[14]. Items of the same size appeared intermittently in a fixed interleaved schedule. Each size occurred once in every seven presentations. Counterbalancing the books between accounts also balanced the timing of their individual prices.

Both eyes remained open. A desk fixture held the paper at 40 cm from the forehead stop, and the reader could keep their usual reading glasses on throughout. The examiner drew back a cover, waited for a single response and closed the cover at eight seconds. The test clock stopped between items. A participant could decline an item by saying pass. Examiners accepted a correction before the cover closed. They did not read any digits back. There was a ten-minute interval away from print between books. We expected some practice in finding the box to survive this interval, which was one reason to randomize the order rather than always testing the restaurant first.

A scorer transcribed the final spoken digits. The separator could be named or omitted and local words for zero were permitted. A reading counted as correct only if all four digits agreed with the key. Silence, a pass, and a reading of the amount above the box each scored zero. A second scorer resolved recordings in which a correction overlapped closure. The primary endpoint was the number of correct amounts in the book, including every scheduled item. We also retained the number of items on which the reader offered no amount, the time until the final spoken digit of a correct response, and the number of disagreements about scoring. Response times were descriptive because they can be observed only for a selected set of successfully completed items[15].

At the end of period two, ten fresh items at log size 0.3 comprised a concealed-account block. Five would be settled for each owner, according to a schedule the cashier had sealed before the visit. The reader knew there were both kinds of bill but did not know which was which. The small-print task was unchanged. This block asked whether a difference remained when the allocation existed on the ledger alone. It did not repeat the exposed comparison at the same point in the visit, and fatigue could alter the average success rate. We did not compare this average directly with the pooled average of the main books.

Statistical analysis

For participant i and account a, let C_{ia} be the sum of the 35 indicators of exact reading. We defined the contrast before comparing any totals. The positive direction favors the reader’s own account.

$$D_i = C_{i,\text{personal}} - C_{i,\text{restaurant}}, \qquad \widehat{\Delta} = \frac{1}{n}\sum_{i=1}^{n} D_i. \tag{1}$$

The unit of replication in equation (1) is the person. We used a two-sided paired t interval with 119 degrees of freedom at full enrollment and tabulated the distribution of D_i to show how well the mean represented individual experience. Because each size contributed five items, a difference of five readings corresponds to one line of the scoring grid. We did not use this conversion to estimate a change in the minimum resolvable angle. In particular the smallest item read correctly by chance can have considerable influence on a threshold derived from that single item[16,17].

A pilot standard deviation of four amounts for the within-person difference led to a planned half-width of approximately 0.72 amounts at 120 complete pairs. Each center

accepted 60 allocations. The protocol called for retaining the first-period scores if a participant left before the second period, with a separate first-period estimate and a report of the missing pair. Unpresented items would be missing for that book, distinct from a pass after presentation. Secondary analyses described accuracy across sizes and by sequence. Book and allocation were crossed by design. We specified no test for equality of carryover effects, which cannot usefully be separated from treatment by sequence in a two-period study[10,18]. For the concealed block we tabulated the two totals and did not specify an equivalence margin.

Results

Of 151 patients approached, 18 could not enter (eight were outside the postoperative window, six could not complete the large-print exercise, and four were needed for further clinical assessment). Of the 133 eligible patients, 13 declined the extra visit time. All 120 allocated patients read both books, took part in the concealed-account block and received the assigned settlement. There were 60 first allocations to each account and 30 of each at each site. No participant crossed to the alternative sequence. Baseline measurements refer to the period before either instruction, as set out in Table 1.

Table 1. Patients at randomization, by first account. Values are counts or medians (interquartile range).

Characteristic	Personal first (60)	Restaurant first (60)
Age, years	71 (66–77)	72 (65–78)
Women	32	34
Weeks since most recent operation	7 (6–9)	7 (5–9)
Previous surgery in the fellow eye	39	36
Wore near correction during testing	49	51
Binocular near letter score, logMAR	0.20 (0.10–0.30)	0.20 (0.10–0.32)
Parma / Aarhus	30 / 30	30 / 30

There were 3,288 exact readings of personal amounts and 2,842 of restaurant amounts in 4,200 presentations per account. The difference was 3.72 amounts per book (95% CI 3.14 to 4.30), or 10.6 percentage points of all presentations. A positive paired difference occurred in 101 people, with eight ties and 11 negative differences. The largest restaurant advantage was five amounts and the largest personal advantage was 13. The full distribution in Figure 1 includes those tails. The distribution is displayed in people. A reader identifying ten more charges on their own account enters the plot once.

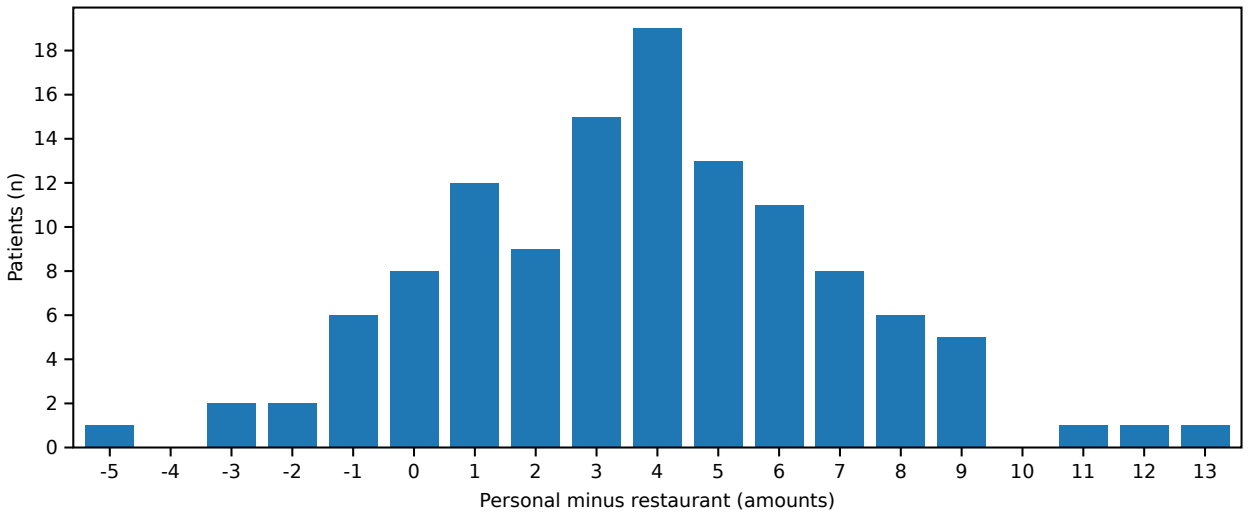


Figure 1. Differences between personal and restaurant scores within patients. Bar heights give the number of patients with that exact difference.

The sequence totals in Table 2 show that patients assigned the personal account first read nearly 28 amounts correctly in their first book. The between-person difference in the first period was 4.65 amounts. Comparing periods within each sequence yielded 3.95 more on the personal account in that sequence and 3.48 more on it in the reverse sequence. Both sequences contribute equally to equation (1). The size of the difference in first periods alone is less precisely determined and includes any chance differences between the two groups of patients.

Table 2. Exact readings by randomized sequence. Each period contains 2,100 presentations in 60 patients.

Assigned first	First book, total	Second book, total	Personal minus restaurant, mean
Personal	1,679	1,442	3.95
Restaurant	1,400	1,609	3.48

At the two largest print sizes almost everybody could read the charge regardless of settlement. The display in Figure 2 uses the actual digit height and includes the smallest size at the left. At log size 0.3 the two accounts yielded 449 and 342 correct amounts. The largest numerical difference was at 0.2 (336 versus 220). At 0.1 the corresponding counts were 229 and 128. Reading fell below 40% even on the personal account at that size. A person who had to recover the cover charge still needed a sufficiently large price.

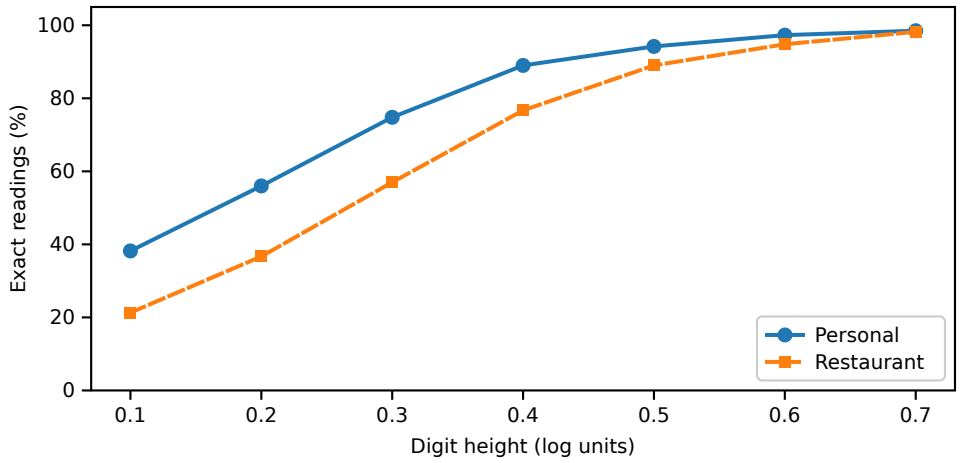


Figure 2. Exact readings by log digit height. Each point represents 600 scheduled amounts per account. Lines connect observed proportions.

We examined omissions separately because all four digits had to be attempted to earn credit. On the restaurant account 648 entries received no amount, compared with 307 on the personal account. The remaining errors also fell, from 710 to 605. A substantial fraction of the additional exact readings could therefore have arisen from making an

attempt at all. These totals do not say that the omitted entries on one account would have been read exactly on the other. Different amounts appeared in the two periods. The breakdown of responses in Figure 3 refers to all scheduled presentations, including amounts that generated no speech.

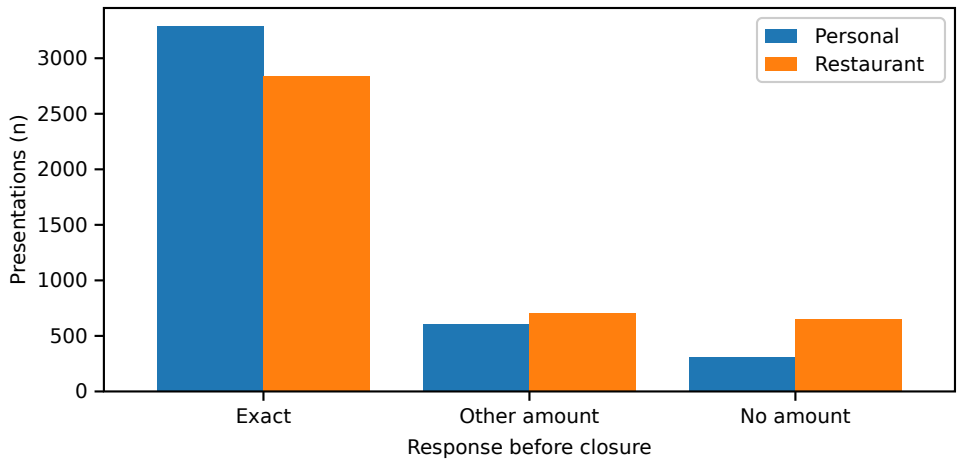


Figure 3. Disposition of the 4,200 presentations in each account, classified by the final response before closure.

Among successful readings, median response time was 4.1 seconds (interquartile range 2.9 to 5.8) for personal amounts and 3.7 (2.6 to 5.0) for restaurant amounts. The additional successful readings on the personal account included many at small sizes. In 37 of 8,400 main presentations the transcript required a second scorer. Of these, 24 concerned a correction, eight concerned a local expression for the decimal part and five concerned an examiner moving the paper at closure. The scorers settled all 37 without inspecting the account. The fixed fee was paid on every visit. No participant asked to stop because of discomfort, and no study-related adverse event was recorded during the visit.

The sealed block contributed another 1,200 presentations. The cashier’s personal column contained 399 exact amounts and the restaurant column 395, out of 600 each. A total of 49 patients read more of the five sealed personal amounts, 46 read more of the restaurant amounts and 25 had equal scores. The two columns had no visible labels to distinguish them. Several readers kept a pencil tally of which column they believed was theirs. A tally of personal and restaurant charges supplied by the participant had no standing in settlement. Staff used the sealed assignment for every refund.

Discussion

The addition of a beneficiary to a disputed line item altered the number of small amounts read at a postoperative visit. The clinical letter score had been recorded before this manipulation. The power of the reading addition was fixed, as were the distance to the page and the examiner’s time limit. Where the amounts were large enough for almost universal success the identity of the creditor had little room to make a difference. At smaller sizes the difference involved both willingness to offer four digits and success in offering the correct ones. Some readers who had given up promptly on the restaurant’s bill spent the full interval on their own.

We chose disputed amounts because their correct reading can have an actual consequence while the print remains identical in design. Arranging for the restaurant to receive a refund imposed a real external beneficiary and avoided a comparison between real payment and an imaginary exercise. This choice has a cost. Participants may have cared about the restaurant, the refreshments bought for other patients or the appearance of doing a good job for the examiner. Any of these could raise the

restaurant score. Conversely, a reader who thought the company should check its own receipts may have declined charges that were entirely legible. The task incorporates these decisions. Pseudophakia provides no reason to expect immunity to them.

The neat conversion of the mean difference to 0.074 log units is a property of the scoring grid. It comes from assigning one fifth of a 0.1 step to each amount at any size. Cross-chart agreement requires evidence about individual differences and their dispersion[14,19]. Two people can have the same score by transcribing quite different distributions of digit sizes. In particular, the opportunity to read another large amount does not let a reader resolve a smaller decimal point. The direction of the conversion is also easy to reverse: more correct responses correspond to a smaller equivalent print size. The actual observations in Figure 2 are the firmer basis for comparing conditions.

These patients had made a trip to a postoperative clinic and could read the large practice example. People whose clinical course kept them in the examining room were missing by design. Our recipients were told exactly who owned each account, at a time when the probability of payment was also explained. In everyday dining the ownership of an item can be uncertain until the bill has circulated around the table. A shared starter, a service charge distributed over several diners, or reimbursement by a third party would leave the credit fraction undetermined. A single account indicator cannot then reproduce the task in Methods. No conclusion about chart performance in those settlement arrangements follows from these observations.

The examiner was aware of which account was open and might have allowed slight differences in the time at closure. The recorded timestamp covered the full exposure;

the five recordings with an ambiguous closure represent the cases where this could affect a correction directly. More subtle cues in posture remain possible. Repeating the study with the examiner behind a partition would remove some cues but would also remove the ordinary social setting in which somebody else presents a bill. A second limitation is that refusal and inability to read both count as zero. This was intentional for an outcome defined by exact recovery of the printed amount. An inference concerning sensory thresholds would require a task that separates refusals from incorrect forced choices.

This study supplies a condition to record when everyday bills are used for assessment. In these data a four-digit amount did not have a single reading score until the expense office had specified whose money it was.

Governance and records

The Valderi Joint Clinical Review Board approved the protocol and both translations (VJ-24-86). Participants gave written consent, including consent to audio scoring. Fixed compensation did not depend on a dispute. Both sites paid research expenses from their clinical measurement budgets. The authors report no financial interest in the accounts. The numerical material needed to reproduce the primary interval is given in the appendix. Recordings of identifiable voices remain with the recruiting center.

Score reconstruction

The sums for Figure 2 are provided in Table 3 because rounded percentages do not reproduce the participant-level interval. Each cell has a denominator of 600; the ratio

differences across cells describe a size effect in this set of books. Adding the seven differences and dividing by 120 recovers the within-person mean. To calculate its standard error one also needs the paired distribution in Figure 1. Its sample standard deviation is 3.205 amounts. Dividing by the square root of 120 gives 0.293, before applying the two-sided critical value. Retaining an independent observation for each amount would use the wrong sampling unit for the trial[20].

Table 3. Correct readings at each print size, from smallest to largest.

Log digit height	Personal	Restaurant	Difference
0.1	229	128	101
0.2	336	220	116
0.3	449	342	107
0.4	534	460	74
0.5	565	534	31
0.6	584	569	15
0.7	591	589	2

For the exposed condition each pair of books was matched on the four positions of the price, although exact matching on all four simultaneously would have repeated the price and compromised the crossover. The match was on the marginal frequency of digits in each position. We printed prices on both sides of the median from each account into every size band. For the concealed block the cashier alone held the five-and-five allocation; the identical seal was opened whether or not an amount had been recovered. These distinctions were retained on the settlement form so that a refund recovered by recognizing the envelope could not acquire the status of a reading.

The operational form also recorded a request to hold the receipt closer. Staff could pause between items to adjust the forehead stop for comfort. They could not bring an

open page nearer the participant. In nine visits the participant took the entire closed book in hand during a pause. Staff returned it to the fixture before showing the next amount. The paired score includes those visits. Dropping the reader who reached for a bill would select on interest in its contents, one of the behaviors the account assignment might reasonably have changed.

Acknowledgments

We thank the clinic cashiers who reconciled the research credits and the patients who read both books.

References

1. Bailey IL, Lovie JE. New design principles for visual acuity letter charts. American Journal of Optometry and Physiological Optics. 1976;53:740–745.
2. Bailey IL, Lovie JE. The design and use of a new near-vision chart. American Journal of Optometry and Physiological Optics. 1980;57:378–387.
3. Ferris FL, Kassoff A, Bresnick GH, Bailey I. New visual acuity charts for clinical research. American Journal of Ophthalmology. 1982;94:91–96.
4. Legge GE, Rubin GS, Pelli DG, Schleske MM. Psychophysics of reading. II. Low vision. Vision Research. 1985;25:253–265.
5. Radner W. Standardization of reading charts: a review of recent developments. Optometry and Vision Science. 2019;96:768–779.
6. Lovie-Kitchin J. Reading with low vision: the impact of research on clinical management. Clinical and Experimental Optometry. 2011;94:121–132.
7. Almari T, Donceli I. Anticipation of prices in a grocery-label near reading test. Valderi Annals of Functional Optics. 2021;11:31–39.
8. Berlen G, Ossvik V. Selection of everyday print samples brought to a postoperative clinic. Trestin Reports on Clinical Measurement. 2022;7:60–67.
9. Dwan K, Li T, Altman DG, Elbourne D. CONSORT 2010 statement: extension to randomised crossover trials. BMJ. 2019;366:l4378.
10. Senn S. Cross-over Trials in Clinical Research. Wiley; 2002.
11. Sarde M, Fiorel A. Localization of an itemized charge in restricted fields of near vision. Valderi Annals of Functional Optics. 2020;10:77–86.
12. Prenn N, Embris J. Digit predictability in till and table receipts. North Quay Transactions in Applied Reading. 2023;4:18–27.
13. Vellami C, Hesden R. Scale-dependent visual cues from nontext elements in printed test materials. Trestin Reports on Clinical Measurement. 2019;4:44–51.

14. Bailey IL, Lovie-Kitchin JE. Visual acuity testing. From the laboratory to the clinic. Vision Research. 2013;90:2–9.
15. Delrac I, Nuvol M. Selection of spoken latency observations in timed reading tests. Vestdal Papers in Physiological Optics. 2024;8:12–21.
16. Wichmann FA, Hill NJ. The psychometric function: I. Fitting, sampling, and goodness of fit. Perception and Psychophysics. 2001;63:1293–1313.
17. Korell S, Andravi R. Score granularity and end-of-chart errors in mixed-digit recognition. Vestdal Papers in Physiological Optics. 2022;6:70–79.
18. Grizzle JE. The two-period change-over design and its use in clinical trials. Biometrics. 1965;21:467–480.
19. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1:307–310.
20. Altman DG. Practical Statistics for Medical Research. Chapman and Hall; 1991.