

The Auckland Optotypes: An open-access pictogram set for measuring recognition acuity

Lisa M. Hamm

School of Optometry and Vision Science, and New Zealand National Eye Centre, University of Auckland, Auckland, New Zealand



Janice P. Yeoman

School of Optometry and Vision Science, and New Zealand National Eye Centre, University of Auckland, Auckland, New Zealand

Nicola Anstice

School of Optometry and Vision Science, and New Zealand National Eye Centre, University of Auckland, Auckland, New Zealand
Optometry and Vision Science, University of Canberra, Canberra, Australia



Steven C. Dakin

School of Optometry and Vision Science, and New Zealand National Eye Centre, University of Auckland, Auckland, New Zealand
UCL Institute of Ophthalmology, University College London, London, UK



When measuring recognition acuity in a research setting, the most widely used symbols are the Early Treatment of Diabetic Retinopathy Study (ETDRS) set of 10 Sloan letters. However, the symbols are not appropriate for patients unfamiliar with letters, and acuity for individual letters is variable. Alternative pictogram sets are available, but are generally comprised of fewer items. We set out to develop an open-access set of 10 pictograms that would elicit more consistent estimates of acuity across items than the ETDRS letters from visually normal adults. We measured monocular acuity for individual uncrowded optotypes within a newly designed set (The Auckland Optotype [TAO]), the ETDRS set, and Landolt Cs. Eleven visually normal adults were assessed on regular and vanishing formats of each set. Inter-optotype reliability and ability to detect subtle differences between participants were assessed using intraclass correlations (ICC) and fractional rank precision (FRP). The TAO vanishing set showed the strongest performance (ICC = 0.97, FRP = 0.90), followed by the other vanishing sets (Sloan ICC = 0.88, FRP = 0.74; Landolt ICC = 0.86, FRP = 0.80). Within the regular format, TAO again outperformed the existing sets (TAO ICC = 0.77, FRP = 0.75; Sloan ICC = 0.65, FRP = 0.64; Landolt ICC = 0.48, FRP = 0.63). For adults with normal

visual acuity, the new optotypes (in both regular and vanishing formats) are more equally legible and sensitive to subtle individual differences than their Sloan counterparts. As this set does not require observers to be able to name Roman letters, and is freely available to use and modify, it may have wide application for measurement of acuity.

Introduction

Determining the smallest letter or shape that a person can reliably identify has been a key component of visual assessment for the last 150 years (Westheimer, 2016). However, clinical measures of recognition acuity (henceforth acuity) can have poor test–retest reliability. Agreement between test and retest ranges from less than ± 1 line¹ (± 0.10 logMAR) to over ± 3 lines on a standard eye chart (95% limits of agreement [Siderov & Tiu, 1999; Rosser, Laidlaw, & Murdoch, 2001; Rosser, Murdoch, & Cousens, 2004]). Such variability is due to factors both extrinsic and intrinsic to the test. Extrinsic factors include how the test is administered (Bailey & Lovie, 1976; Strong & Woo, 1985; Bailey & Lovie-

Citation: Hamm, L. M., Yeoman, J. P., Anstice, N., & Dakin, S. C. (2018). The Auckland Optotypes: An open-access pictogram set for measuring recognition acuity. *Journal of Vision*, 18(3):13, 1–15, <https://doi.org/10.1167/18.3.13>.

<https://doi.org/10.1167/18.3.13>

Received August 26, 2017; published March 23, 2018

ISSN 1534-7362 Copyright 2018 The Authors

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.



Downloaded From: <http://jov.arvojournals.org/> on 04/19/2018

Kitchin, 2013), such as termination and scoring criteria (Vanden Bosch & Wall, 1997; Carkeet, 2001; Shah, Dakin, Whitaker, & Anderson, 2014), and characteristics of the test group such as their cognitive function (Elyashiv, Shabtai, & Belkin, 2014) and visual status (Rosser et al., 2004). When test administration is well-controlled but participants are recruited through clinics, measures can be 95% accurate within ± 0.1 logMAR across ages (assessed using coefficient of repeatability [CoR]; Beck et al., 2003) and ± 0.125 logMAR in children (CoR; Holmes et al., 2001). This can be reduced further to as low as ± 0.06 logMAR in laboratory settings with psychophysically experienced adult observers (CoR; Arditi & Cagenello, 1993). Remaining variability is largely attributable to intrinsic aspects of the test and/or variability in human performance. In this paper, we consider one intrinsic and fundamental aspect of acuity tests, the symbols used.

Within a research setting the most widely used symbols for acuity assessment are the 10 Early Treatment of Diabetic Retinopathy Study (ETDRS) Sloan letters (Bailey & Lovie-Kitchin, 2013). Two desirable features of these symbols are (a) their stroke width is consistently 1/5 of their overall size, and (b) their 1:1 aspect ratio. Two less desirable features are that (a) Roman letters are unfamiliar to some observers, and (b) some letters are easier to identify than others at a given size (Bennett, 1965; Ferris, Freidlin, Kassoff, Green, & Milton, 1993; Alexander, Xie, & Derlacki, 1997). The organization of ETDRS letter charts attempts to compensate for variation in letter difficulty (Bailey & Lovie, 1976; Strong & Woo, 1985) by mixing harder and easier items within a line. However, this option is not available when symbols are presented singly (for example in some electronic tests [Beck et al., 2003; Yamada et al., 2015]) and is less effective when scoring is by letter rather than by line.

The recommended optotypes for observers who do not know their letters (for example, many preschool children) are a truncated set of Sloan letters (HOTV) or a set of pictograms known as the Lea symbols (Hyvarinen, Nasanen, & Laurinen, 1980; Cotter et al., 2015; Donahue & Baker, 2016). Individual Lea symbols were designed to be similarly legible to one another, have consistent stroke to bounding box ratios (1:7) and close-to-uniform 1:1 aspect ratios. The HOTV set of Sloan letters have the advantage of ease of comparison to ETDRS optotypes, while yielding reliable acuity results for young children (Hered, Murphy, & Clancy, 1997; Holmes et al., 2001; Cyert, 2004, 2010; Cotter et al., 2015; Yamada et al., 2015). Both HOTV and Lea sets contain only four symbols because fewer alternatives reduce cognitive load and are considered easier for children. The downside of 4 (rather than 10) alternative forced choice (AFC) judgments is that they limit the

information gained from each decision, and so reduce confidence in estimates of acuity made over a fixed number of trials (Arditi & Cagenello, 1993; Carkeet, 2001). Other pictogram optotype sets have been developed with more than four shapes. For example, Patti pics (Singman, Matta, Tian, & Silbert, 2015) include five symbols that have the same stroke-to-bounding-box ratio as Sloan letters (facilitating comparison of acuity measures). Other sets include more alternatives such as Allen figures, Amsterdam pictures (Engin et al., 2014), and Kay pictures (Kay, 1983). However, developing larger sets of unique, simple shapes that are equally legible/confusable is difficult. Consequently many sets include items with variable stroke width and/or irregular aspect ratios. These attributes make it difficult to maintain consistency in the placement and size of crowding elements (Cyert, 2010; Lalor, Formankiewicz, & Waugh, 2016), which is a significant disadvantage given the importance of assessing vulnerability to crowding in amblyopia (Stuart & Burian, 1962; Levi & Klein, 1985; Levi, Hariharan, & Klein, 2002; Greenwood et al., 2012). Finally, acuity levels elicited by individual optotypes in these larger groups are variable (Amsterdam Picture [Engin et al., 2014]; Kay pictures [Lalor et al., 2016]; for overall summaries, see Candy, Mishoulam, Nosofsky, & Dobson, 2011; Anstice & Thompson, 2014; Anstice et al., 2017).

Some of these challenges have led to recent interest in redesigning picture optotypes, as seen in the new version of the Kay pictures (Milling, Newsham, Tidbury, O'Connor, & Kay, 2016). Although an encouraging development, these symbols are only available commercially, which limits their access by clinicians and researchers. Alternatively, there are advantages to developing new optotypes openly within the wider vision research community. This gives others the opportunity to assess different applications of the symbols, to conduct independent assessment and validation, and even to modify the symbols themselves. This is possible with Sloan letters, which are downloadable in a vector graphic format (which is modifiable; see <http://psych.nyu.edu/pelli/software.html>).

One example of innovation with Sloan letters is the development of a “vanishing” or pseudo-high pass version (Howland, Ginsburg, & Campbell, 1978). Here, optotype strokes are made of a central white or black band, surrounded on either side by a finer band of opposite contrast polarity. When the strokes of such optotypes cannot be resolved, the optotype becomes indistinguishable from the gray background and vanishes. Acuity measurements made with vanishing optotypes are less dependent on the number of alternatives, exhibit better test–retest reliability compared to regular optotypes (Shah, Dakin, Redmond, & Anderson, 2011), and reduce errors within acuity

measures in children (Fariza, Kronheim, Medina, & Katsumi, 1990). This is because these letters minimize the low spatial frequency information that supports letter identification beyond our resolution limit for strokes so forcing detection and recognition to occur at the same stimulus size (Frisen, 1986; Fariza et al., 1990; Adoh, Woodhouse, & Oduwaiye, 1992; Shah et al., 2011; Shah, Dakin, & Anderson, 2012). The most widespread application of vanishing optotypes is the Cardiff Acuity Test (Adoh et al., 1992). In this test a vanishing picture is printed in one of two alternative locations, and a preferential looking paradigm used to determine threshold. The Cardiff Acuity Test is useful for estimating acuity in very young children and those with cognitive difficulties (Adoh et al., 1992). However, results do not align well with recognition acuity tests using regular optotypes (Paudel et al., 2017), perhaps due to perception of vanishing compared to regular optotypes, or perhaps because of the testing protocol. It would therefore be useful to compare measures of acuity made with vanishing and regular versions of identical optotypes with the same test protocol, as has been done with Sloan letters (Shah et al., 2011; Shah et al., 2012; Shah, Anderson, Tufail, Egan, & Dakin, 2013; Shah et al., 2014). Finally, we note that optotype features such as internal junctions and acute angles can prevent uniform vanishing of pseudo-high pass symbols. Therefore, it is feasible to optimize the benefits of the vanishing optotypes by including additional constraints to promote vanishing in the design phase.

In this paper, we describe the development of a new set of 10 pictogram optotypes. We ensured that each item had consistent stroke width, had a 1:1 aspect ratio, was fully enclosed with no internal lines, had limited acute angles, and was sufficiently unique to minimize ambiguity in identification (either by naming or matching). Here, we investigated whether our new optotype set could improve reliability of individual acuity estimates within a set. To accomplish this, we compared individual, isolated optotypes within the newly designed set (The Auckland Optotypes [TAO]) to two others: (a) the ETDRS Sloan letter set and (b) Landolt Cs presented at four orientations (C, C, C, C). We assessed both regular and vanishing formats of each. We used the ETDRS Sloan letter set because it is the set most commonly used (in a research setting) with adults and children 7 and older (for example, Beck et al., 2003), and is also the largest (10-item) well standardized set. We used a Landolt C set because items are simple rotations of a single symbol, so maximizing inter-item similarity. We compared the optotype sets in terms of internal reliability (precision), and ability to capture small differences between participants (sensitivity) using intraclass correlations ([ICC]; McGraw & Wong, 1996) and fractional rank precision ([FRP]; Dorr et al., 2017). This paper

describes the development of an optotype set for use with participants who may be unfamiliar with the Roman alphabet (such as children). Note however that, in order to reduce the impact of confounding variables, we carried out the testing with healthy adults viewing isolated optotypes, presented using standard psychophysical procedures. Further work, some of which is ongoing (Hamm et al., 2016), will be required to determine how best the new optotypes might be used to evaluate vision in children, individuals with cognitive impairment, and patients with visual loss.

Methods

Development of optotypes

We used a vector-graphics editing program (Illustrator CS9; Adobe Systems Inc., San Jose, CA) to first generate a large set of candidate pictograms. We used thinner strokes than Sloan optotypes (stroke to bounding box ratio of 1:8.23 rather than 1:5) to allow us to generate a wider range of identifiable items. We then used the “uniqueness” and anticipated nameability of items to reduce this set to 20 items. Custom MATLAB (MathWorks, Natick, MA) scripts were developed to estimate various image properties of the optotypes. These properties were: *area* (proportion of the bounding box occupied by the optotype), *perimetric complexity* (outline perimeter-length squared, divided by the product of the area and 4π [Watson, 2012])² and *mean pairwise-overlap* (the sum of the *centered* pairwise overlap with each remaining optotype in the set, divided by number of pairs). Each is illustrated in Figure 1 (upper middle panel), which shows the image statistics used in Phase 1 of the design. We used these measures to further refine the set down to 10 items with similar image properties.

Next, we used an iterative process to further refine the set. Specifically, we (1) conducted pilot psychophysics experiments, (2) replaced items or adjusted item shape (to deal with outliers), and (3) reran the image analysis described above. This process was repeated until we could not further improve the balance of the set (see below). Steps 1 and 2 are expanded on below.

Pilot experiments included at least five adult participants assessed using the interleaved staircase paradigm described under the Procedure section. We primarily used this pilot work to assess how consistent individual optotype acuity measures were across the set, but we also took this opportunity to elicit informal feedback from participants not captured by the image statistics or threshold estimations. For example, some participants mentioned certain features provided distinctive local cues to pictogram identity; others stated

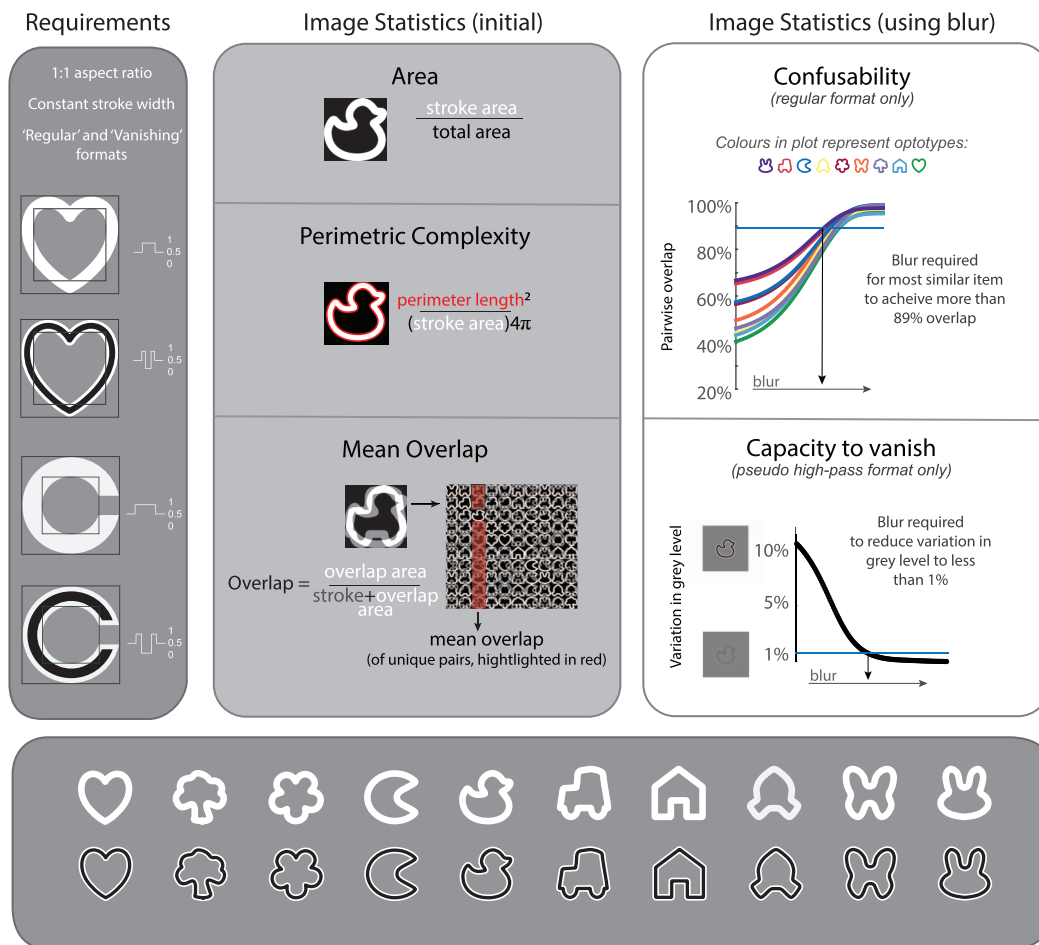





Figure 1. Description of image properties used in optotype development. The final set, in both regular and vanishing formats, is displayed in the lower panel (♥ heart, 🌳 tree, 🌸 flower, 🌙 moon, 🦆 duck, 🚗 car, 🏠 house, 🚀 rocket, 🦋 butterfly, and 🐰 rabbit).

that some items were not sufficiently unique, making identification and naming ambiguous. Based on acuity thresholds we made adjustments to the set to further improve optotype balance and identification. We then returned to MATLAB to recompute image properties of the adjusted set. Optotypes identified as having unique image properties (defined as falling more than 2 *SDs* away from the set mean) were adjusted prior to returning to pilot psychophysics. Early in the development cycle, outlier items were replaced with a different candidate optotype. Later in the cycle, changes were limited to minor adjustments to the shape of individual items. This iterative process was halted when interoptotype variability plateaued. The plateau did not correspond to perfect balance. Rather, we reached a point at which each new change either unbalanced the set or compromised ease of item identification.

When we could not further reduce interitem variability (in terms of the elicited acuity thresholds within a set), we assessed the impact of simulated *blur* on both (a) optotype confusability, and (b) capacity for pseudo high-pass optotypes to vanish. Each is described

graphically in Figure 1 (upper right panel), which illustrates the images statistics used in Phase 2 of the design. To quantify *confusability*, we estimated, on an item-by-item basis, the level of blur required for a given optotype to exceed a predefined (threshold) level of pairwise overlap with another item. More similar items need less blur to exceed overlap threshold (see Supplementary Figure S1). The cut-off value used was not critical and other cut-offs provided similar results. To quantify “capacity for vanishing,” we estimated the blur level that caused the standard deviation of the image gray levels (RMS contrast) to fall below a predetermined threshold. Again, the specific threshold was not critical, we used 1%. This procedure mainly led to minor adjustment of the corners and curves within optotypes to facilitate vanishing. The outcome of this process is pictured in the lower panel of Figure 1; two formats of each pictogram were produced, a full stroke regular symbol and a split stroke vanishing format.³

Upon completion, although the acuity elicited by each optotype fell within 2 *SDs* of the mean across the set, some optotypes remained *potential* outliers. This included the 🦋 (having the largest area, most com-

plexity, and generally eliciting lower thresholds in pilot work), the  (having the highest mean pairwise-overlap and eliciting higher thresholds), the  (having the lowest simulated threshold in the blur and confusability analysis and low thresholds in pilot work), and the  (having the smallest area, lowest complexity, and least mean overlap).

Participants

The project adhered to the tenets of the Declaration of Helsinki and was approved by the University of Auckland Human Participants Ethics Committee. All participants provided informed consent prior to screening for eligibility. Each had vision better than or equal to 0.0 logMAR (measured with ETDRS Sloan letters on the Medmont AT20-R chart (<http://www.medmont.com/products/at20p-visual-acuity-tester/>) with no reported history of visual disorders. Eleven adults, aged 20–35 years, participated in, and successfully completed, this study.











Prior to acuity testing all participants underwent noncycloplegic subjective refraction. Subjective refraction was checked on a Medmont AT20-R chart using standard clinical procedures (plus to first blur and Jackson Crossed Cylinder determination of astigmatism). Habitual refraction was worn if subjective refraction was within 0.50 Dioptres (D). If habitual refraction was out of date, or if the participant was a contact lens wearer, appropriate spectacle correction was provided in trial frames. Only the right eye was tested in all cases.

Apparatus

Testing was conducted at 4 m in a well-lit (minimum 500 Lux) clinical testing room. We ran tests on a Microsoft Surface Pro 3 (Microsoft, Redmond, Washington, USA) computer fitted with a matte screen cover. The LCD display has a native resolution of $2,160 \times 1,440$ pixels running at 60 Hz. Experiments ran under the MATLAB environment (MathWorks) and our code incorporated elements of the Psychtoolbox (psychtoolbox.org; Brainard, 1997; Pelli, 1997; Kleiner, Brainard, Pelli, Ingling, Murray, & Broussard, 2007). Participants used a hand-held labeled keyboard to input responses.

Stimuli

The ETDRS set consisted of the 10 Sloan letters (**C**, **D**, **H**, **K**, **N**, **O**, **R**, **S**, **V**, **Z**), the Landolt Cs consisted of a Sloan **C** presented at one of four orientations (**C**,

⌵, **⌴**, **⌶**). The Sloan font (Pelli, Robson, & Wilkins, 1988) used was downloaded from <http://psych.nyu.edu/pelli/software.html>. TAO (Figure 1) consisted of 10 items (, , , , , , , , , ). Each of the three sets of optotypes were generated in two formats (regular and vanishing), leading to six distinct set/conditions. The strokes of our regular optotypes appeared white (~ 300 cd/m²) on a gray (~ 150 cd/m²) background. Strokes of vanishing optotypes had a black center stroke (~ 1 cd/m²) surrounded by two half-width white bands, displayed on a gray background. The polarity of the regular optotypes was selected to approximate the mean luminance of the vanishing stimuli, facilitating comparison.

Procedure














Participants were familiarized with the new shapes prior to testing, and each participant completed a 10-min training program to become familiar with the testing protocol. On a single trial, a single optotype appearing in the center of a gray screen for 500 ms, immediately followed by a screen displaying response options (all of the possible optotypes in set). The selected optotype flashed green if the participant was correct and red if incorrect. A Bayesian adaptive staircase procedure, QUEST (Watson & Pelli, 1983), controlled the optotype size (over ~ 40 trials) to determine the minimum size supporting 69% recognition performance. We term this minimum size the *threshold size*. QUEST was given a prior threshold estimate of 0.2 logMAR above the set mean threshold of the first few participants (the first few participants started with generally higher priors). We assigned an estimated standard deviation of 0.3 logMAR. The lapse rate was set to 0.01 and the guess-rate (γ) set to 1 divided by number of alternatives (4 or 10).

Note that we could not assess acuity for a given n -item optotype set using a standard n -AFC procedure because this yields the mean threshold across the set and not for individual optotypes. Thus, in our procedure each staircase assessed the threshold for a single optotype so that staircases had to be interleaved in a run (containing only items of its set and condition) to generate alternatives. Each run was repeated three times for both standard and vanishing conditions of all three sets (with order of testing randomized). Together, each participant completed approximately 5,760 trials over approximately 10 hr, spread across several sessions.

The methodological choice of *effective* n -AFC via interleaving single optotype staircases was necessary, but had the potential to underestimate threshold. For example, should an observer correctly identify the letter **S** several times early on in a run, QUEST could reduce

letter size to well below the item's identification threshold, so that the unusual appearance of the stimulus (as a smaller, illegible character) might be sufficient to identify it. We minimized the probability of such runaway staircases by using a pseudorandom (rather than full random) procedure to interleave staircases. Specifically, each experimental run was divided into sequences of either eight (Landolt C) or 20 (Sloan and TAO) trials, where each contained two full sets of all alternatives (in random order). This procedure made it extremely unlikely the same optotype would appear more than twice in a row (e.g., for a 10-item optotype set the probability of three similar optotypes appearing sequentially was only 1.2%).

Analysis

We defined a runaway staircase as one containing less than three errors yielding a threshold more than 2.5 SDs from the set mean for condition. Using these criteria, the regular format of TAO had 12 out of 330 impacted staircases, or 3.6% (six , two , one , , , and ) , the Sloan regular format set had 14 out of 330 or 4.2% of staircases eliminated (13 **V** and one **K**) and the Landolt Cs regular format had five out of 132 or 3.8% runaway staircases (two , two , and one ) . Within vanishing conditions, two out of 330 staircases (0.6%) within TAO set were eliminated (one  and ) , Sloan vanishing had eight out of 330 instances or 2.4% (six **V**, one **H** and **S**), and Landolt Cs vanishing had one out of 132 () runaway staircase or 0.8%. Note that certain optotypes were more likely to elicit a runaway staircase. In particular the **V** (in both regular and vanishing formats) and to a lesser extent the  (only in the regular format). Removing runaway staircases from further analysis provided a conservative estimate of differences between optotype acuity thresholds, so we focused on this analysis, but a summary of raw data is also provided.

For each participant, responses for each optotype across the three runs were pooled (discounting stated runaway staircases) and fitted with a cumulative normal psychometric function using the Palamedes toolbox. An ideal optotype set would have high precision (highly reliable acuity estimates between optotypes within a set), but also be sensitive to the subtle acuity differences between participants. An ICC provides an estimate capturing aspects of each, but is prone to distortion by outliers. Recently, a new statistic for measuring test reliability has been proposed (Dorr et al., 2017) called fractional rank precision (FRP). FRP employs an information retrieval approach and evaluates a test by quantifying how identifiable a participant is from their set of test scores. For a test to produce values that identify the participants, scores

must both vary across participants and be consistent within participant. FRP scores how well a test can identify participants from 0.5 (chance) to 1.0 (perfect identification). We used both FRP and ICC measures to evaluate each set/condition combination. In all cases logMAR acuity thresholds were calculated based on the stroke width.

Results

Optotype acuity in the context of set and condition

The mean threshold for each set was lowest for TAO (−0.36 logMAR), followed by the Sloan letters (−0.31 logMAR) and then the Landolt Cs (−0.28 logMAR). For the vanishing format, TAO were very similar to the Landolt Cs (Landolt Cs: −0.31 logMAR and TAO: −0.31 logMAR), with Sloan letters elicited slightly better acuity results (Sloan: −0.35 logMAR). Note that we should be cautious comparing these results to standard measures of acuity (where 0.0 logMAR is a notional norm) given the differences in protocol. In particular, TAO (like all new optotype sets) would need to be calibrated to produce acuity estimates aligned with clinical standards (Bailey & Lovie-Kitchin, 2013).

Figure 2 plots estimated visual acuity measured with each optotype within the set and condition to which it belongs. Sets are displayed in columns (left shows results from TAO, middle from Sloan, and right from Landolt Cs) and conditions in rows (the upper row shows regular and the lower row vanishing formats). The acuity threshold from each observer for each optotype is shown as a small colored symbol (where color denotes observer). The group mean (across individuals, for one optotype) is marked with the corresponding pictogram. Shaded regions represent 95% confidence intervals (CIs) for a single optotype across all observers. ICCs are reported for each set/condition combination. Higher ICC values demonstrate more consistent acuity estimates between optotypes; the ICC value denotes the proportion of variance attributable to differences between subjects.

Indeed, not all participants had the same acuity thresholds despite each scoring at least 0.0 logMAR on our eligibility screening task (measured on the Medmont acuity system, comprised of the 10 ETDRS Sloan letters). For example, in Figure 2 the participants coded with purple and blue tended to have higher thresholds (poorer acuity), while the participants coded with orange and red tended to have lower thresholds (better acuity). In Figure 3 we re-express the data from Figure 2, but with participants rather than optotypes on the x-

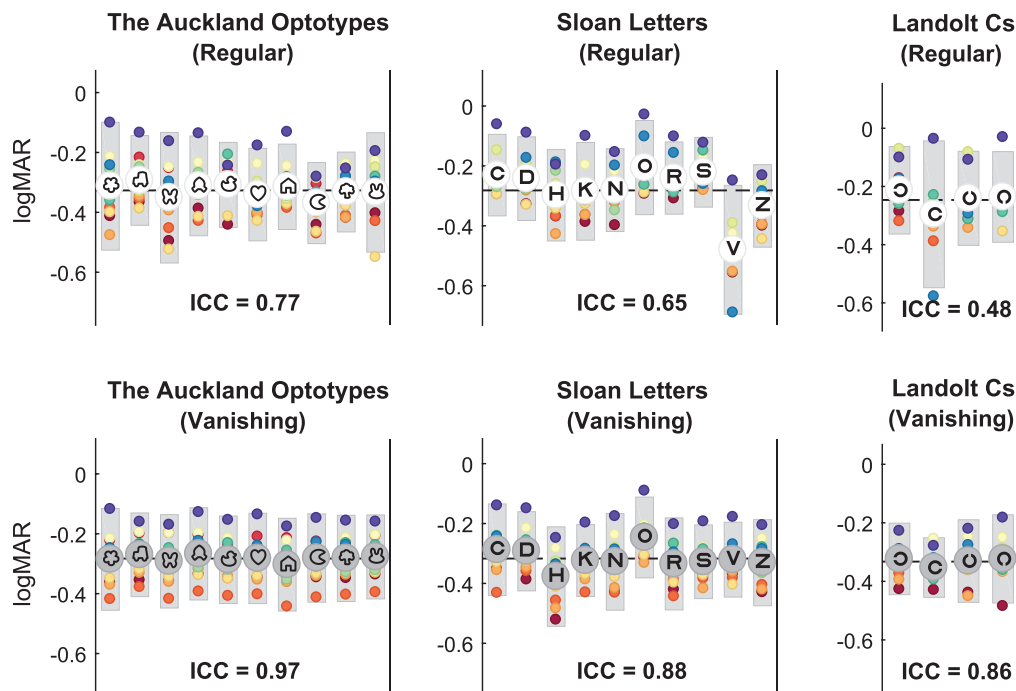


Figure 2. Acuity outcomes by optotype. The y-axis displays visual acuity thresholds in logMAR. Each colored circle represents one participant's results; each optotype symbol represents mean performance for that optotype with the CI shown in gray. ICC is reported for each set/condition combination.

axis. Optotypes are represented by their symbols, with 95% CIs represented by shading in the corresponding participant's color. Colored CIs shows how reliable a participant was with all the optotypes within a set. The F statistic from the ICC is provided. Higher F values

indicated better sensitivity, and reflect the visual impression from Figure 3 that some sets (particularly TAO vanishing format) allow more refined differentiation between participants with small visual acuity differences.

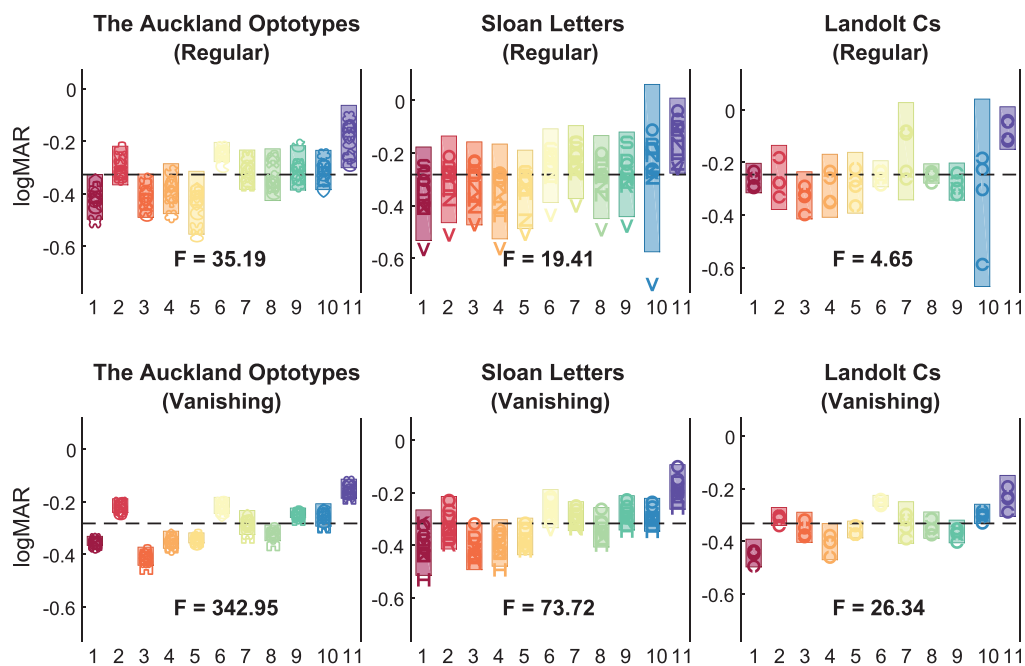


Figure 3. Acuity outcomes by participant. The x-axis denotes Participants 1 to 11, and the y-axis displays visual acuity thresholds in logMAR. Each colored symbol represents the participant's results for the pictured optotype. Shaded areas represent 95% CIs for each participant for the particular set/condition. The intraclass correlation F statistic is reported for each set/condition combination.

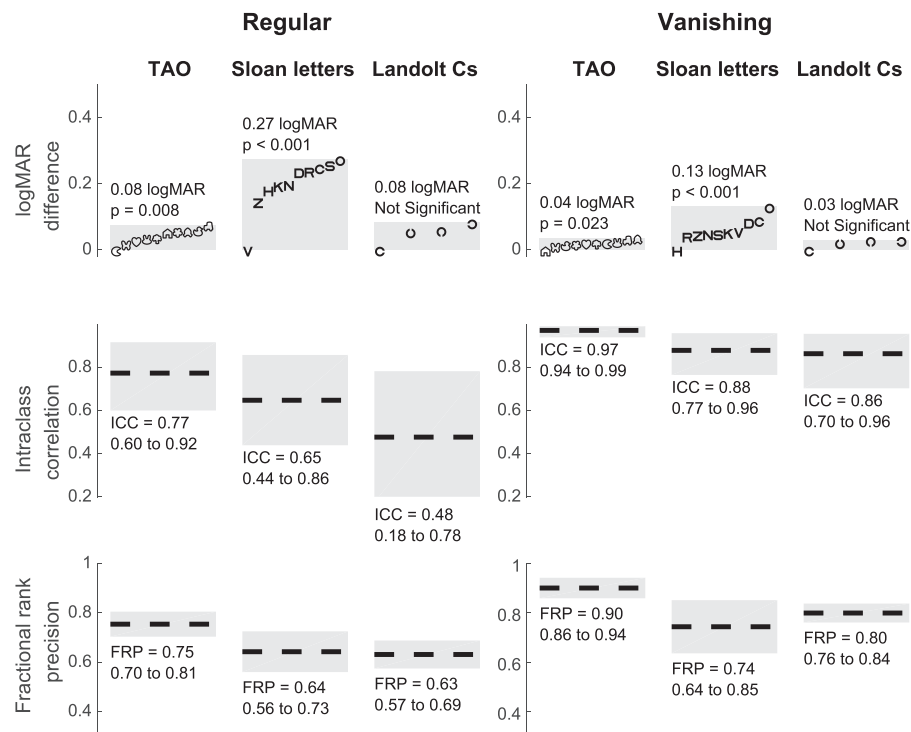


Figure 4. Summary of comparison between sets and conditions. Three methods of comparison are shown. In the first row, the difference between the easiest and the most difficult optotype within each set is highlighted (including whether there was a group effect of optotype within the rmANOVA). In the second row, the ICC is represented by a dotted line and text, with upper and lower limits shaded. Finally, fractional rank precision is summarized in the bottom row by dotted lines and text, with upper and lower limits shaded.

In Figure 4 we summarize three different measures of set reliability. The top row is the most straightforward, highlighting simply the range of mean acuity thresholds. On this measure, the vanishing formats of Landolt Cs and TAO both show very little variation from the easiest to the most difficult items (0.03 and 0.04 logMAR, respectively, or less than half a line on a standard eye chart). By contrast, the regular Sloan letters have a range of 0.28 logMAR from acuity estimates generated from the **V** and that from the **O** (almost three lines on a standard eye chart). The p value denotes level of significance from a repeated measures analysis of variance (ANOVA) for each set/condition, with less than 0.05 recorded as not significant. ICC estimates are presented in Row 2, summarized from Figure 2 and additionally including upper and lower bounds of the estimates. A higher mean (dotted line) suggests a more reliable optotype set. The bottom row shows the results for FRP. This measure incorporates both precision and sensitivity, whereby a score of 1 represents perfect test–retest identification. On all measures TAO outperform Sloan letters and Landolt Cs, and vanishing formats outperform their counterpart regular optotype set. When the same measures are calculated on the data set including all runaway staircases (clamped at a lower limit or left uncon-

strained) range and FRP maintain TAO as more reliable than Sloan in both regular and vanishing formats. ICC, however, becomes less informative with increasing outlier data, and the results vary depending on lower limits.

Error analysis

Figure 5 shows a proportional summary of responses made across all participants and trials. The displayed optotype is shown on the y -axis and reported optotype on the x -axis. Common responses are highlighted in yellow. In each matrix the yellow diagonal cells represents correct responses, while off-diagonal yellow cells represent common errors.

Common errors could be due to optotype similarity (for example the **D** and the **O** have high pairwise overlap and this combination is a common error highlighted in Figure 5), or due to a bias towards a particular item (perhaps **O** is a favored response). Luce's Choice Model aids in teasing apart similarity from bias. Using this strategy, Figure 6 displays individual biases and Figure 7 the quasisymmetric similarity matrices collapsed across all participants.

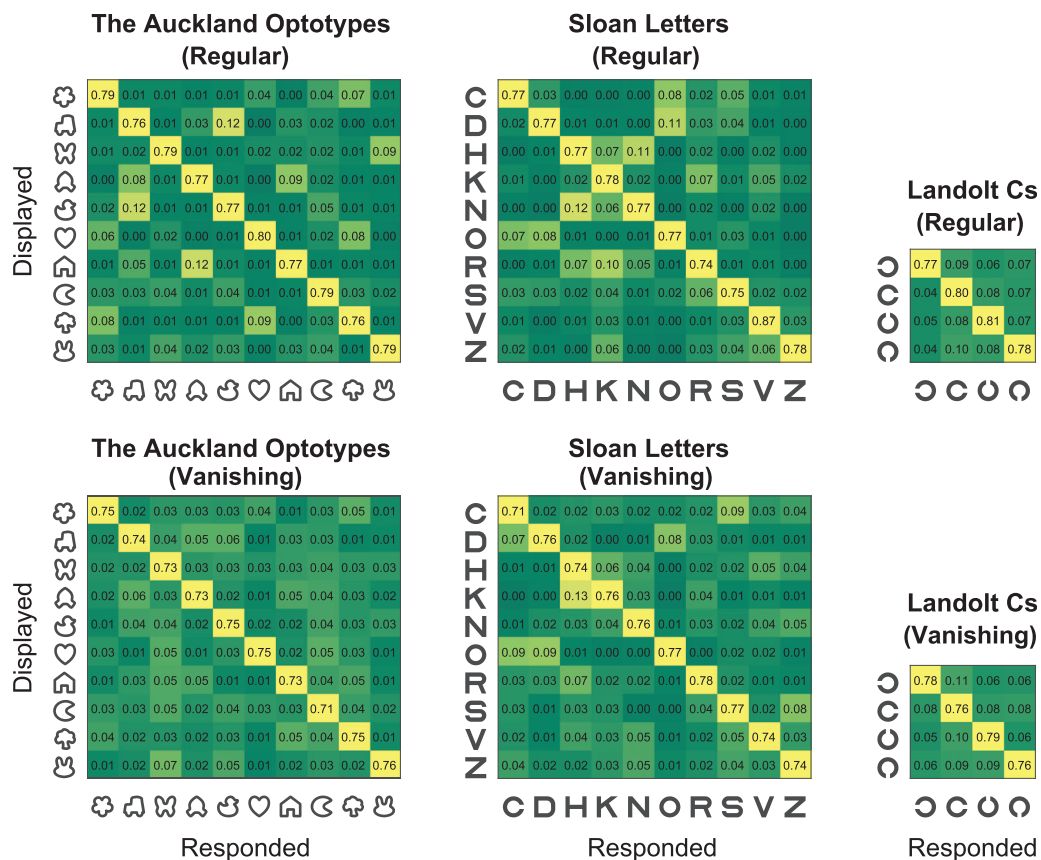


Figure 5. Confusion matrices. The y -axis displays the optotype shown to the participant, and the x -axis the responses. The upper row shows regular formats and the lower row results from vanishing formats. Numbers in each matrix represent the proportion of decisions in response to the displayed optotype, with common responses highlighted in yellow.

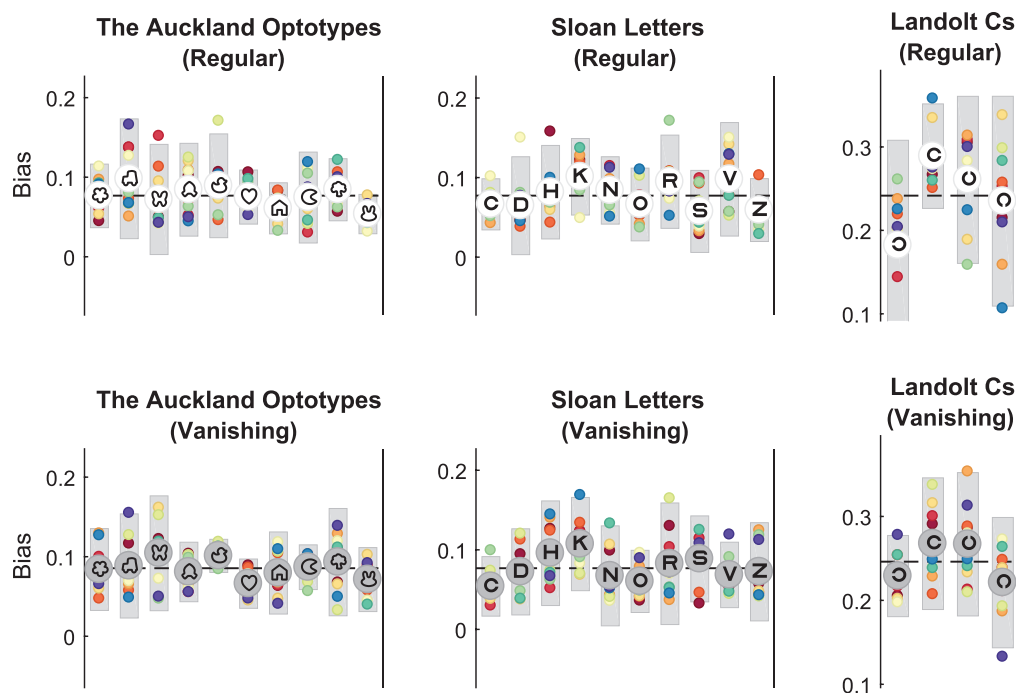


Figure 6. Individual bias. Luce's Choice model was used to extract individual bias to respond with a particular optotype. Participant colors are consistent with that presented in Figures 2 and 3.

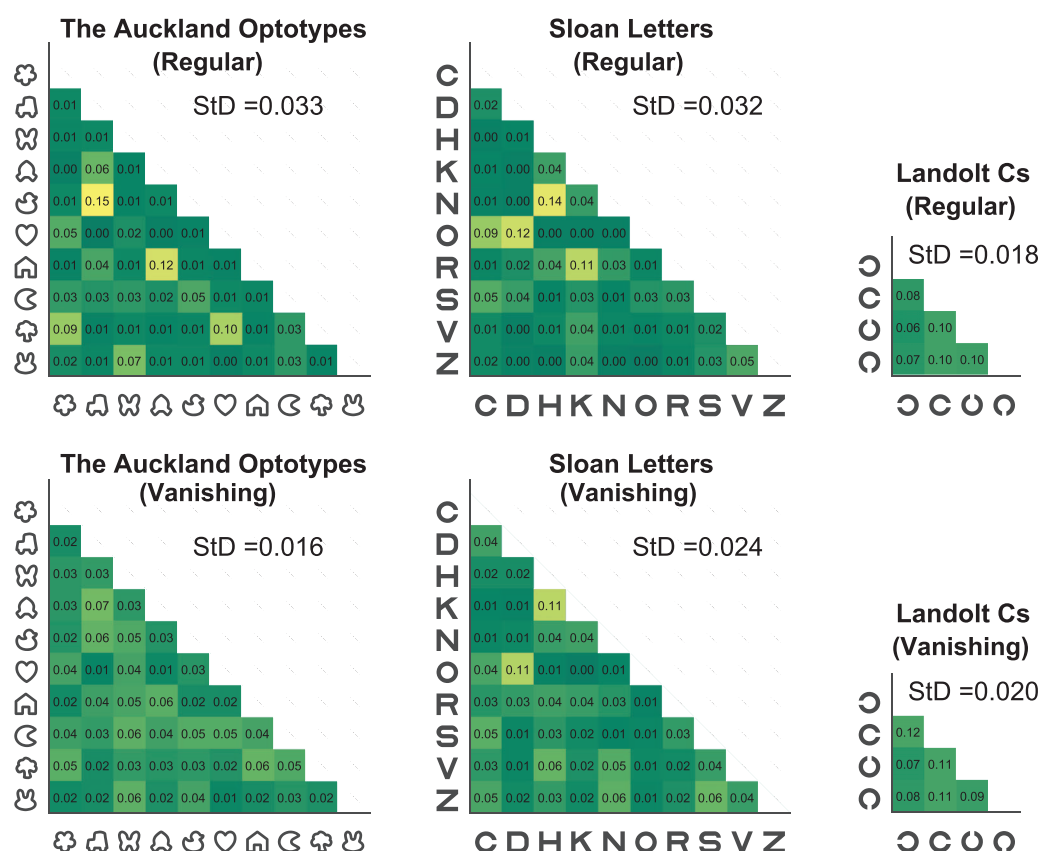


Figure 7. Quasisymmetric similarity. Luce's Choice model was used to calculate similarity individually. This data was collapsed across participants and displayed as a triangle matrix. Yellow cells suggest the two corresponding optotypes are perceived to be similar. Variance is reported with StD in the top corner of each matrix. As in Figure 5, regular formats are displayed in the top row, and vanishing on the bottom.

Variation in response bias appeared consistent between sets (regular: TAO = 0.028, Sloan = 0.030; vanishing: TAO = 0.027, Sloan = 0.032). There was a trend towards more bias for Landolt Cs (regular = 0.066, vanishing = 0.042), with the rightward facing gap (the orientation corresponding to the letter) favored.

Comparing the yellow off-diagonal cells in Figures 5 and 7, it appears that similarity is driving the pattern of errors, rather than bias. For regular optotype formats, we anticipate errors are more likely for optotypes sharing low spatial frequency shape cues, as the majority of trials are displayed near threshold where high spatial frequencies are unresolvable. Indeed, we find a significant correlation between pairwise overlap (one of our image statistics) and similarity (as shown in Figure 7; TAO regular: $R^2 = 0.59$, $p < 0.001$; Sloan regular: $R^2 = 0.58$, $p < 0.001$), and this correlation is stronger when the overlap is assessed with blurred images driven by low spatial frequencies (TAO regular: $R^2 = 0.64$, $p < 0.001$; Sloan regular: $R^2 = 0.70$, $p < 0.001$). Since vanishing optotypes contain minimal low spatial frequency information, we expect errors to be more equally distributed across all possible options. In other words, we expect the standard deviation (RMS

contrast) of the confusion matrix to decrease as optotypes vanished more completely. We find a pattern consistent with this hypothesis; after eliminating correct responses the standard deviation of the confusion matrices ranged from 0.028 (for both the regular formats of TAO and Sloan letters) to 0.024 for vanishing Sloan letter and 0.014 for TAO vanishing formats, a pattern maintained in the Luce similarity matrices, and presented in Figure 7 (standard deviations reported on each subplot). The most common error in the vanishing sets, highlighted by a yellow patch in the bottom middle panel of Figure 5, is mistaking a **K** for an **H**. This was consistent with an observation made during our design phase that the **K** was the optotype most resistant to vanishing in the blur simulations.

Discussion

Acuity is a critical measure of visual function for preschool vision screening (Simons, 1996; Cotter et al., 2015), clinical eye care (Donahue & Baker, 2016), and

research (Tsirlin, Colpa, Goltz, & Wong, 2015; Guo et al., 2016). Current acuity measurements are limited by test–retest variability (Beck et al., 2003; Rozhkova, Podugolnikova, & Vasiljeva, 2005), patient’s familiarity with letters, the ease or difficulty with which individual letters may be recognized (Strong & Woo, 1985; Ferris et al., 1993; Alexander et al., 1997), and for picture optotypes, copyright limitations. In this study, we report the introduction of an open-access set of unique pictograms. We endeavored to generate a set of 10 optotypes that would address several shortcomings in existing optotype sets of a similar number of alternatives. Specifically, we designed items to have the following features: consistent stroke width, a 1:1 aspect ratio, fully enclosed shapes with no internal lines, and limited acute angles. Given these constraints, we investigated whether it was feasible to design a set of 10 optotypes in which the variability of acuity estimates within set was between that of the ETDRS Sloan letters (upper boundary; Strong & Woo, 1985; Ferris et al., 1993; Alexander et al., 1997) and Landolt Cs (lower boundary). We used psychophysical protocols to answer the question in visually normal adult participants. We found that the precision and discriminatory power of the new TAO set was better than the ETDRS set in both regular and vanishing format, and elicited results similar to Landolt Cs. As such, this new optotype set is potentially quite useful to identify small differences in recognition acuity, at least between adults with normal visual acuity. This was particularly the case for the newly designed vanishing set.

More work is required to assess whether these benefits translate to higher test–retest reliability. Indeed, the utility of this optotype now needs to be assessed across a wide range of cognitive abilities, visual acuities, cultural backgrounds, and ages. Further, given the importance of assessing susceptibility to crowding in the diagnosis of amblyopia (Stuart & Burian, 1962; Levi & Klein, 1985; Levi et al., 2002; Greenwood et al., 2012), these symbols would need to be tested in a crowded format, and compared to crowded versions of current symbols optotypes. Three of our design constraints (consistent stroke width, a 1:1 aspect ratio, and fully enclosed shapes with no internal lines) make the new stimuli particularly useful for studies of crowding.

The role of vanishing optotypes is potentially of further interest; vanishing optotypes have lower test–retest variability (Shah et al., 2014) and may be particularly sensitive for detecting macular pathology (Shah et al., 2016). Starting with enclosed shapes and intentionally limiting acute angles and internal lines reduced the presence of high contrast regions that can arise in vanishing Sloan letters (e.g., at line endings or junctions), further reducing interoptotype variability. This feature of the proposed optotype set is an

additional advantage over other well designed pictogram sets (for example Lea symbols [Hyvarinen et al., 1980] and the newly designed Kay pictures [Milling et al., 2016]), adding a level of flexibility to the type of testing which can be done. For example, since vanishing optotypes may be helpful for monitoring visual loss arising from macular degeneration (Shah et al., 2016), having a set of 10 vanishing optotypes that do not rely on the familiarity with Roman letters could be useful in a culturally diverse clinical setting. Whether there is an appropriate clinical application of vanishing optotypes for children is less well understood. Vanishing symbols appear practical for children; early work suggests vanishing optotypes may lead to more accurate acuity measures in children (Fariza et al., 1990), and within the Cardiff Acuity Test, vanishing symbols are effective for testing infants and those with cognitive impairment (Adoh et al., 1992). However, whether some childhood conditions differentially impact vanishing or regular acuity thresholds has not been explored as it has in adults (Shah et al., 2013; Shah et al., 2016). Since TAO allows direct comparison between regular and vanishing formats, it could be a useful tool to further explore clinical applications for vanishing optotypes in children.

Our work supports previous studies which have found differences in legibility within the Sloan ETDRS letter set (Strong & Woo, 1985; Ferris et al., 1993; Alexander et al., 1997; Reich & Bedell, 2000). Chart design compensates for this, with each line containing items of different individual difficulty but similar mean difficulty (Strong & Woo, 1985; Ferris et al., 1993; Ferris & Bailey, 1996), but interoptotype differences become more problematic when testing a single optotype. As observers approach the (average) threshold size for recognition, previous studies indicate that rounded letters (**C**, **D**, and **O**) are harder to recognize than letters with straight lines (**N**, **V**, and **Z** Strong & Woo, 1985; Ferris et al., 1993; Alexander et al., 1997). Errors also follow predictable trends, with circular letters such as **C**, **D**, and **O** being easily confused, as well as those with strong vertical components such as **N** / **H** and **R** / **K** (Reich & Bedell, 2000). Our results for legibility and errors are generally consistent with this literature. However, in previous work the **V** has not been identified as an outlier to the extent it was here. Strong and Woo (1985) and Reich and Bedell (2000) showed data from their own group as well as Sloan’s original work, suggesting that **Z** is the easiest optotype to recognize, followed closely by **N** and **V**. Alexander et al. (1997) had a protocol more similar to our own, and two of the three participants in their study found the **V** the easiest optotype to recognize. It is possible that differences in methodology between our own (interleaved QUEST staircase procedure) and earlier studies may have exaggerated differences between optotypes.

Nevertheless, our main goal was to compare precision between sets, in which case, exaggerated differences would reinforce the finding that the new set yields very reliable estimates between optotypes. Although equal legibility with an optotype set is thought to be advantageous (Sloan, 1951; Bailey & Lovie, 1976; Anstice & Thompson, 2014), there is some evidence that improving interoptotype variability does not improve test–retest reliability (Arditi & Cagenello, 1993; Raasch, Bailey, & Bellmore, 1998; Shah et al., 2014). This is an important area of future research for both multiple line chart based, and single optotype, adaptive acuity measures.

For adults with poor reading skills, or for whom language prevents use of standardized symbols, this group of 10 pictograms allows equity with the current 10-alternative standard for acuity testing. Paediatric optotype tests tend to use smaller numbers of alternatives (typically four [Hyvarinen et al., 1980; Holmes et al., 2001], six [Milling et al., 2016], or eight [Kay, 1983]). Although the increased number of symbol alternatives in TAO has the potential to increase test efficiency, whether this translates to a benefit when testing children remains to be seen, particularly as increased alternatives can increase indecision and cognitive load. Intuitive nameability reduces cognitive load and should help children deal with the higher number of alternatives. Although nameability was a motivator for our design, it was not formally tested in this study. Our current work with children in New Zealand and Tonga suggests our optotypes are nameable, with the ♀ eliciting the widest *variety* of names (alternatives to ‘rocket’ include ‘shield’, ‘hat’, and even ‘campfire’; Hamm et al., 2016). The optotypes are unique enough that the creativity in naming does not appear a limiting factor (similar to the heart or apple item within the Lea symbols). To promote engagement, we have developed simple animations illustrating possible symbol names, which may be useful when introducing new observers to a short acuity test using the symbols. More work across ages and cultures is needed to assess appropriateness of the stimuli, and to explore their sensitivity and specificity for detecting clinically significant visual disorders such as amblyopia, strabismus, and refractive error in both adults and children.

Further work on this set could take a variety of forms. More detailed models of the optical, neural, and cognitive limits on optotype identification (Watson & Ahumada, 2012) will be useful to further refine the proposed pictogram set. For example, careful selection of a subset for a particular purpose (for example, to maximize the impact of astigmatism on performance) could be a valuable expansion of this work. Further analysis of the degree to which various image statistics predict identification accuracy and error type would

also be valuable. Another important step would be to investigate the impact of crowding on recognition of each optotype individually. Such questions are best addressed through a combination of modeling, psychophysics with adult participants, and use of clinically relevant methodology (including testing across various ages and pathologies).

We have made the optotypes and related material freely available (see footnote 3 for download information) and would encourage other groups to assess their appropriateness for various applications. Our hope is that a freely available optotype set, which is developed through a transparent and collaborative process, will prove beneficial in a variety of settings.

Keywords: acuity, optotypes, recognition, children’s vision

Acknowledgments

We are grateful to Cure Kids New Zealand (project 3707907) and to the Robert Leidl Trust for supporting this project. We thank Jonathan Albert for assistance with data collection and Matt Wilson for his help with font conversion and animation.

Commercial relationships: none.

Corresponding author: Lisa M. Hamm.

Email: l.hamm@auckland.ac.nz.

Address: School of Optometry and Vision Science, University of Auckland, Auckland, New Zealand.

Footnotes

¹ Where 1 line is $0.1 \times \log_{10}$ of the minimum angle of resolution (logMAR).

² We calculated perimeter from the vector format in Illustrator rather than estimating it from a binary digital image as described in Watson (2012).

³ Symbols and additional resources can be freely accessed from <https://github.com/dakinlab/OpenOptotypes>.

References

- Adoh, T. O., Woodhouse, J. M., & Oduwaiye, K. A. (1992). The Cardiff test: A new visual acuity test for toddlers and children with intellectual impairment. A preliminary report. *Optometry and Vision Science*, 69(6), 427–432.
- Alexander, K. R., Xie, W., & Derlacki, D. J. (1997).

- Visual acuity and contrast sensitivity for individual Sloan letters. *Vision Research*, 37(6), 813–819.
- Anstice, N. S., Jacobs, R. J., Simkin, S. K., Thomson, M., Thompson, B., & Collins, A. V. (2017). Do picture-based charts overestimate visual acuity? Comparison of Kay Pictures, Lea Symbols, HOTV and Keeler logMAR charts with Sloan letters in adults and children. *PLoS One*, 12(2), e0170839.
- Anstice, N. S., & Thompson, B. (2014). The measurement of visual acuity in children: An evidence-based update. *Clinical and Experimental Optometry*, 97(1), 3–11.
- Arditi, A., & Cagenello, R. (1993). On the statistical reliability of letter-chart visual acuity measurements [Abstract]. *Investigative Ophthalmology and Visual Science*, 34(1), 120–129.
- Bailey, I. L., & Lovie, J. E. (1976). New design principles for visual acuity letter charts. *American Journal of Optometry and Physiological Optics*, 53(11), 740–745.
- Bailey, I. L., & Lovie-Kitchin, J. E. (2013). Visual acuity testing. From the laboratory to the clinic. *Vision Research*, 90, 2–9.
- Beck, R. W., Moke, P. S., Turpin, A. H., Ferris, F. L., III, SanGiovanni, J. P., Johnson, C. A., ... Kraker, R. T. (2003). A computerized method of visual acuity testing. *American Journal of Ophthalmology*, 135(2), 194–205.
- Bennett, A. G. (1965). Ophthalmic test types. A review of previous work and discussions on some controversial questions. *The British Journal of Physiological Optics*, 22(4), 238–271.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433–436.
- Candy, T., Mishoulam, S. R., Nosofsky, R. M., & Dobson, V. (2011). Adult discrimination performance for pediatric acuity test optotypes. *Investigative Ophthalmology and Visual Science*, 52(7), 4307–4313, <https://doi.org/10.1167/iovs.10-6391>.
- Carkeet, A. (2001). Modeling logMAR visual acuity scores: Effects of termination rules and alternative forced-choice options. *Optometry and Vision Science*, 78(7), 529–538.
- Cotter, S. A., Cyert, L. A., Miller, J. M., Quinn, G. E., Russ, S. A., Block, S. S., ... Wallace, D. K. (2015). Vision screening for children 36 to 72 Months: Recommended practices. *Optometry and Vision Science*, 92(1), 6–16.
- Cyert, L. (2004). Preschool visual acuity screening with HOTV and Lea symbols: Testability and between-test agreement. *Optometry and Vision Science*, 81(9), 678–683.
- Cyert, L. (2010). Effect of age using lea symbols or HOTV for preschool vision screening. *Optometry and Vision Science*, 87(2), 87–95.
- Donahue, S. P., & Baker, C. N. (2016). Procedures for the evaluation of the visual system by pediatricians. *Pediatrics*, 137(1), 1–9.
- Dorr, M., Elze, T., Hui, W., Lu, Z., Bex, P. J., & Lesmes, L. A. (in press). New precision metrics for contrast sensitivity testing. *IEEE Journal of Biomedical and Health Informatics*, <https://doi.org/10.1109/JBHI.2017.2708745>.
- Elyashiv, S. M., Shabtai, E. L., & Belkin, M. (2014). Correlation between visual acuity and cognitive functions. *The British Journal of Ophthalmology*, 98(1), 129–132.
- Engin, Ö., Despret, D. D. G., van der Meulen-Schot, H. M., Romers, A., Slot, X., Sang, M. T. F., ... Simonsz, H. J. (2014). Comparison of optotypes of Amsterdam Picture Chart with those of Tumbling-E, LEA Symbols, ETDRS, and Landolt-C in non-amblyopic and amblyopic patients. *Graefes Archive for Clinical and Experimental Ophthalmology*, 252(12), 2013–2020.
- Fariza, E., Kronheim, J., Medina, A., & Katsumi, O. (1990). Testing visual acuity of children using vanishing optotypes. *Japanese Journal of Ophthalmology*, 34(3), 314–319.
- Ferris F. L., III, & Bailey, I. (1996). Standardizing the measurement of visual acuity for clinical research studies: Guidelines from the Eye Care Technology Forum. *Ophthalmology*, 103(1), 181–182.
- Ferris F. L., III, Freidlin, V., Kassoff, A., Green, S. B., & Milton, R. C. (1993). Relative letter and position difficulty on visual acuity charts from the Early Treatment Diabetic Retinopathy Study. *American Journal of Ophthalmology*, 116(6), 735–740.
- Frisen, L. (1986). Vanishing optotypes. New type of acuity test letters. *Archives of Ophthalmology*, 104(8), 1194–1198.
- Greenwood, J. A., Tailor, V. K., Sloper, J. J., Simmers, A. J., Bex, P. J., & Dakin, S. C. (2012). Visual acuity, crowding, and stereo-vision are linked in children with and without amblyopia. *Investigative Ophthalmology & Vision Science*, 53(12), 7655–7665.
- Guo, C. X., Babu, R. J., Black, J. M., Bobier, W. R., Lam, C. S. Y., Dai, S., ... Uren, S. L. (2016). Binocular treatment of amblyopia using video-games (BRAVO): Study protocol for a randomised controlled trial. *Trials*, 17(1), 1–9.
- Hamm, L. M., Langridge, F., Yeoman, J., Grant, C., Fakakovikaetau, T., Anstice, N. S. ... Dakin, S. C. (2016). Childhood vision screening in Tonga

- [Abstract]. *International Congress of Paediatrics*. Vancouver, Canada.
- Hered, R. W., Murphy, S., & Clancy, M. (1997). Comparison of the HOTV and lea symbols charts for preschool vision screening. *Journal of Pediatric Ophthalmology and Strabismus*, 34(1), 24–28.
- Holmes, J. M., Beck, R. W., Repka, M. X., Leske, D. A., Kraker, R. T., Blair, R. C., ... Hertle, R. W. (2001). The amblyopia treatment study visual acuity testing protocol. *Archives in Ophthalmology*, 119, 1345–1353.
- Howland, B., Ginsburg, A., & Campbell, F. (1978). High pass spatial frequency letters as clinical optotypes. *Vision Research*, 18(8), 1063–1066.
- Hyvarinen, L., Nasanen, R., & Laurinen, P. (1980). New visual acuity test for pre-school children. *Acta Ophthalmologica*, 58(4), 507–511.
- Kay, H. (1983). New method of assessing visual acuity with pictures. *British Journal of Ophthalmology*, 67(2), 131–133.
- Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3? *Perception*, 36(14), 1–16.
- Lalor, S. J. H., Formankiewicz, M. A., & Waugh, S. J. (2016). Crowding and visual acuity measured in adults using paediatric test letters, pictures and symbols. *Vision Research*, 121, 31–38.
- Levi, D. M., Hariharan, S., & Klein, S. A. (2002). Suppressing and facilitatory spatial interactions in amblyopic vision. *Vision Research*, 42(11), 1379–1394.
- Levi, D. M., & Klein, S. A. (1985). Vernier acuity, crowding and amblyopia. *Vision Research*, 25, 979–991.
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1), 30–46.
- Milling, A., Newsham, D., Tidbury, L., O'Connor, A., & Kay, H. (2016). The redevelopment of the Kay picture test of visual acuity. *British and Irish Orthoptic Journal*, 206(13), 12–21.
- Paudel, N., Jacobs, R. J., Sloan, R., Denny, S., Shea, K., Thompson, B., & Anstice, N. (2017). Effect of simulated refractive error on adult visual acuity for paediatric tests. *Ophthalmic and Physiological Optics*, 37(4), 521–530.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442.
- Pelli, D. G., Robson, J. G., & Wilkins, A. J. (1988). The design of a new letter chart for measuring contrast sensitivity. *Clinical Vision Sciences*, 2(3), 187–199.
- Raasch, T. W., Bailey, I. L., & Bellmore, M. A. (1998). Repeatability of visual acuity measurement. *Optometry and Vision Science*, 75(5), 342–348.
- Reich, L. N., & Bedell, H. E. (2000). Relative legibility and confusions of letter acuity targets in the peripheral and central retina. *Optometry and Vision Science*, 77(5), 270–275.
- Rosser, D. A., Laidlaw, D. A. H., & Murdoch, I. E. (2001). The development of a reduced logMAR visual acuity chart for use in routine clinical practice. *British Journal of Ophthalmology*, 85(4), 432–436.
- Rosser, D. A., Murdoch, I. E., & Cousens, S. N. (2004). The effect of optical defocus on the test-retest variability of visual acuity measurements [Abstract]. *Investigative Ophthalmology and Visual Science*, 45(4), 1076–1079.
- Rozhkova, G. I., Podugolnikova, T. A., & Vasiljeva, N. N. (2005). Visual acuity in 5-7-year-old children: Individual variability and dependence on observation distance. *Ophthalmic and Physiological Optics*, 25(1), 66–80.
- Shah, N., Anderson, R., Tufail, A., Egan, C., & Dakin, S. (2013). Visual acuity loss in patients with AMD, measured using a vanishing optotype letter chart [Abstract]. *Investigative Ophthalmology and Visual Science*, 54(15), 5021.
- Shah, N., Dakin, S. C., & Anderson, R. S. (2012). Effect of optical defocus on detection and recognition of vanishing optotype letters in the fovea and periphery. *Investigative Ophthalmology and Visual Science*, 53(11), 7063–7070, <https://doi.org/10.1167/iovs.12-9864>.
- Shah, N., Dakin, S. C., Dobinson, S., Tufail, A., Egan, C. A., & Anderson, R. S. (2016). Visual acuity loss in patients with age-related macular degeneration measured using a novel high-pass letter chart. *British Journal of Ophthalmology*, 100(10), 1346–1352.
- Shah, N., Dakin, S. C., Redmond, T., & Anderson, R. S. (2011). Vanishing optotype acuity: Repeatability and effect of the number of alternatives. *Ophthalmic and Physiological Optics*, 31(1), 17–22.
- Shah, N., Dakin, S. C., Whitaker, H. L., & Anderson, R. S. (2014). Effect of scoring and termination rules on test-retest variability of a novel high-pass letter acuity chart. *Investigative Ophthalmology and Visual Science*, 55(3), 1386–1392, <https://doi.org/10.1167/iovs.13-13340>.
- Siderov, J., & Tiu, A. L. (1999). Variability of measurements of visual acuity in a large eye clinic. *Acta Ophthalmologica Scandinavica*, 77(6), 673–676.

- Simons, K. (1996). Preschool vision screening: Rationale, methodology and outcome. *Survey of Ophthalmology*, 41(1), 3–30.
- Singman, E. L., Matta, N. S., Tian, J., & Silbert, D. I. (2015). Comparing visual acuity measured by lea symbols and patti pics. *American Orthoptic Journal*, 65(1), 94–98.
- Sloan, L. L. (1951). Measurement of visual acuity: A critical review. *AMA Archives of Ophthalmology*, 45(6), 704–725.
- Strong, G., & Woo, G. C. (1985). A distance visual acuity chart incorporating some new design features. *Archives of Ophthalmology*, 103(1), 44–46.
- Stuart, J. A., & Burian, H. M. (1962). A study of separation difficulty: Its relationship to visual acuity in normal and amblyopic eyes. *American Journal of Ophthalmology*, 53(3), 471–477.
- Tsirlin, I., Colpa, L., Goltz, H. C., & Wong, A. M. F. (2015). Behavioral training as new treatment for adult amblyopia: A meta-analysis and systematic review. *Investigative Ophthalmology and Visual Science*, 56(6), 4061–4075, <https://doi.org/10.1167/iovs.15-16583>.
- Vanden Bosch, M. E., & Wall, M. (1997). Visual acuity scored by the letter-by-letter or probit methods has lower retest variability than the line assignment method. *Eye (London)*, 11(Pt 3), 411–417.
- Watson, A. B. (2012). Perimetric complexity of binary digital images: Notes on calculation and relation to visual complexity. *Mathematica Journal*, 14, 1–40.
- Watson, A. B., & Ahumada, A. J. (2012). Modeling acuity for optotypes varying in complexity. *Journal of Vision*, 12(10):19, <https://doi.org/10.1167/12.10.19>. [PubMed] [Article]
- Watson, A. B., & Pelli, D. G. (1983). Quest: A Bayesian adaptive psychometric method. *Perception and Psychophysics*, 33(2), 113–120.
- Westheimer, G. (2016). Optotype recognition under degradation: Comparison of size, contrast, blur, noise and contour-perturbation effects. *Clinical and Experimental Optometry*, 99(1), 66–72.
- Yamada, T., Hatt, S. R., Leske, D. A., Moke, P. S., Parrucci, N. L., Reese, J. J., ... Holmes, J. M. (2015). A new computer-based pediatric vision-screening test. *Journal of AAPOS*, 19(2), 157–162.