

# Refinement and preliminary evaluation of two tablet-based tests of real-world visual function

Pete R Jones<sup>1</sup> , Iris Tigchelaar<sup>2,3,4</sup> , Giorgia Demaria<sup>2,5</sup> , Iain Wilson<sup>6,7</sup> , Wei Bi<sup>1</sup>,  
Deanna J Taylor<sup>1</sup>  and David P Crabb<sup>1</sup> 

<sup>1</sup>Division of Optometry and Visual Science, School of Health Sciences, City, University of London, London, UK, <sup>2</sup>Department of Ophthalmology, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands, <sup>3</sup>Ocusweep, <sup>4</sup>Doctoral Program in Clinical Research, University of Turku and Turku University Hospital, Turku, Finland, <sup>5</sup>Graduate School of Medical Sciences (Research School of Behavioral and Cognitive Neurosciences), University of Groningen, Groningen, the Netherlands, <sup>6</sup>Nuffield Laboratory of Ophthalmology, Nuffield Department of Clinical Neurosciences, University of Oxford, and <sup>7</sup>Oxford Eye Hospital, Oxford University Hospitals NHS Foundation Trust, Oxford, UK

**Citation information:** Jones PR, Tigchelaar I, Demaria G, Wilson I, Bi W, Taylor DJ, & Crabb DP. Refinement and preliminary evaluation of two tablet-based tests of real-world visual function. *Ophthalmic Physiol Opt* 2020; 40: 35–46. <https://doi.org/10.1111/opo.12658>

**Keywords:** face discrimination, real world, tablets, visual function, visual search

**Correspondence:** Pete R Jones  
E-mail address: peter.jones@city.ac.uk

Received: 27 August 2019; Accepted: 9 November 2019

**Author contributions:** PRJ: involved in study design; data analysis and interpretation; and drafting and critically revising the manuscript. IT and GD: involved in data acquisition. IW: involved in study design and data acquisition. WB: involved in study design. DJT: involved in study design and critically revising the manuscript. DPC: involved in study conception and design; and critically revising the manuscript.

## Abstract

**Purpose:** To describe, refine, evaluate, and provide normative control data for two freely available tablet-based tests of real-world visual function, using a cohort of young, normally-sighted adults.

**Methods:** Fifty young (18–40 years), normally-sighted adults completed tablet-based assessments of (1) face discrimination and (2) visual search. Each test was performed twice, to assess test-retest repeatability. *Post-hoc* analyses were performed to determine the number of trials required to obtain stable estimates of performance. Distributions were fitted to the normative data to determine the 99% population-boundary for normally sighted observers. Participants were also asked to rate their comprehension of each test.

**Results:** Both tests provided stable estimates in around 20 trials (~1–4 min), with only a further reduction of 14%–17% in the 95% Coefficient of Repeatability (CoR<sub>95</sub>) when an additional 40 trials were included. When using only ~20 trials: median durations for the first run of each test were 191 s (*Faces*) and 51 s (*Search*); test-retest CoR<sub>95</sub> were 0.27 *d* (*Faces*) and 0.84 s (*Search*); and normative 99% population-limits were 3.50 *d* (*Faces*) and 3.1 s (*Search*). No participants exhibited any difficulties completing either test (100% completion rate), and ratings of task-understanding were high (*Faces*: 9.6 out of 10; *Search*: 9.7 out of 10).

**Conclusions:** This preliminary assessment indicated that both tablet-based tests are able to provide simple, quick, and easy-to-administer measures of real-world visual function in normally-sighted young adults. Further work is required to assess their accuracy and utility in older people and individuals with visual impairment. Potential applications are discussed, including their use in clinic waiting rooms, and as an objective complement to Patient Reported Outcome Measures (PROMs).

## Introduction

Traditional clinical measures of basic visual function, such as visual acuity (VA), do not necessarily reflect the patient's experience, or the impact of vision loss on patients' lives.<sup>1</sup> As a result, patient-reported outcome measures (PROMs) are often used as secondary outcome measures (and sometimes as primary outcome measures) in ophthalmic clinical

trials.<sup>2–5</sup> However, discrepancies have been found between self-reports and actual performance on real world tasks.<sup>6–8</sup> Moreover, perceived problems with visual function might be influenced by other factors, such as an individual's lifestyle or personality.<sup>9,10</sup> Methods for directly assessing performance on everyday visually-guided tasks may be more appropriate for assessing the impact of vision loss. Furthermore there is substantial clinical interest in developing

more 'patient friendly' tests that can quickly and easily assess real-world visual function.<sup>11–15</sup>

Our previous work suggests that face discrimination and visual search are two tasks that are particularly important to patients, and are often impaired in people with glaucoma<sup>16–21</sup> or age-related macular degeneration.<sup>22–24</sup> Currently, however, the equipment required for these tests is bulky, and the tests themselves often relatively time-consuming, making them appropriate only for use in research. Rapid, tablet-based versions of these real-world measures would be more clinically applicable. Tablet tests are particularly attractive as they are inexpensive, and because patients in the waiting room can potentially complete them while they wait to be seen – thereby minimising any burden to patients and staff. Indeed, the idea of using tablet-based activities to more productively utilise the time patients spend in waiting areas is a growing area in other medical disciplines. For example, tablets are being increasingly used in health-care to collect questionnaire/PROM data,<sup>25,26</sup> as an educational tool,<sup>27–30</sup> or for functional testing.<sup>31</sup> However, this concept remains relatively unexplored in ophthalmology.

In this paper, we describe two rapid, tablet-based tests for assessing face discrimination and visual search, both of which we have made freely available online (see *Methods*). In the present work, we piloted these tests on a large number of young, normally-sighted 'control' participants. This allowed us to refine the tests, assess usability, and to establish a normative database prior to conducting further studies with patient populations. The objectives of this study were to: (1) measure completion rates, test-retest reliability, test-durations, and ease-of-use for two, novel, tablet-based tests of visual function; (2) refine the tests on the basis of these measurements, and; (3) establish a normative database of expected scores for a young, normally-sighted population.

## Methods

### Overview

Normally-sighted young adults performed two tablet tests designed to assess real-world visual function. One test (*Faces*) measured participants' ability to discriminate between four human faces ('spot the odd one out'). The other test (*Search*) measured participants' ability to locate a particular object in a crowded scene ('find the matching object'). Each test was performed twice, in order to assess within-visit (intrasession) test-retest repeatability. Note that within-visit variability is more likely to reflect the inherent fluctuations due to measurement error ('intrinsic noise'), whereas between-visit variability may also include true vision fluctuations, and so may be greater. The order

of the tests was interleaved, ABAB, with the starting test determined randomly for each participant. Each test was set to run for a fixed — relatively large — number of trials, allowing us to determine, *post hoc*, how many trials were actually required in order to obtain a statistically stable measure of performance. After testing, participants were also asked to rate the ease and clarity of each test, while before testing measures of basic vision (acuity, contrast sensitivity) and cognition (digit recall) were also taken, for comparison.

### Participants

Participants were 50 adults aged 18–40 years (32 female; mean [standard deviation] age: 25 [5] years) with normal or correct-to-normal vision. All participants completed both tests twice. However, for the *Faces* test, data from only 30 individuals are reported. The other 20 participants performed variants of the *Faces* test (e.g., different adaptive algorithm parameters) that were found to substantially less efficient than the final version reported here.

Normal vision was defined as no history of eye disease, and (1) binocular best-corrected letter acuity  $\leq 0.2$  logMAR (tested with an Early Treatment Diabetic Retinopathy Study [ETDRS] chart); (2) binocular best-corrected Pelli-Robson contrast sensitivity  $\geq 1.5$  logCS (tested with Pelli-Robson chart); and (3) a passing score on the 38-plate Ishihara pseudoisochromatic test (Handaya, Tokyo, Japan, 2011 edition). An additional five individuals were recruited (total  $N = 55$ ), but were excluded from the study as they failed to meet all three criteria.

All screening was conducted in a well-lit room by authors IT and GD. Letter acuity and contrast sensitivity charts were presented in a standard lightbox, were scored letter-by-letter,<sup>32</sup> and terminated after more than 50% incorrect responses on a single row. Measures of contrast sensitivity and colour vision loss were included since both tablet tests involved real-world stimuli containing a range of spatial frequencies and contrast levels, and since the *Search* task further included significant chromaticity cues. For completeness, it would also have been desirable to screen for near visual acuity. Its absence as an inclusion criterion in the present study is unlikely to have been problematic given the present cohort of young adults, but reduced near visual acuity may be a substantial confounding factor for older users.

Participants were recruited via advertisements placed around City, University of London. The study was approved by the Ethics Committee for the School of Health Sciences, City, University of London (#ETH1819-0532), and was carried out in accordance with the Declaration of Helsinki. Written informed consent was obtained from all

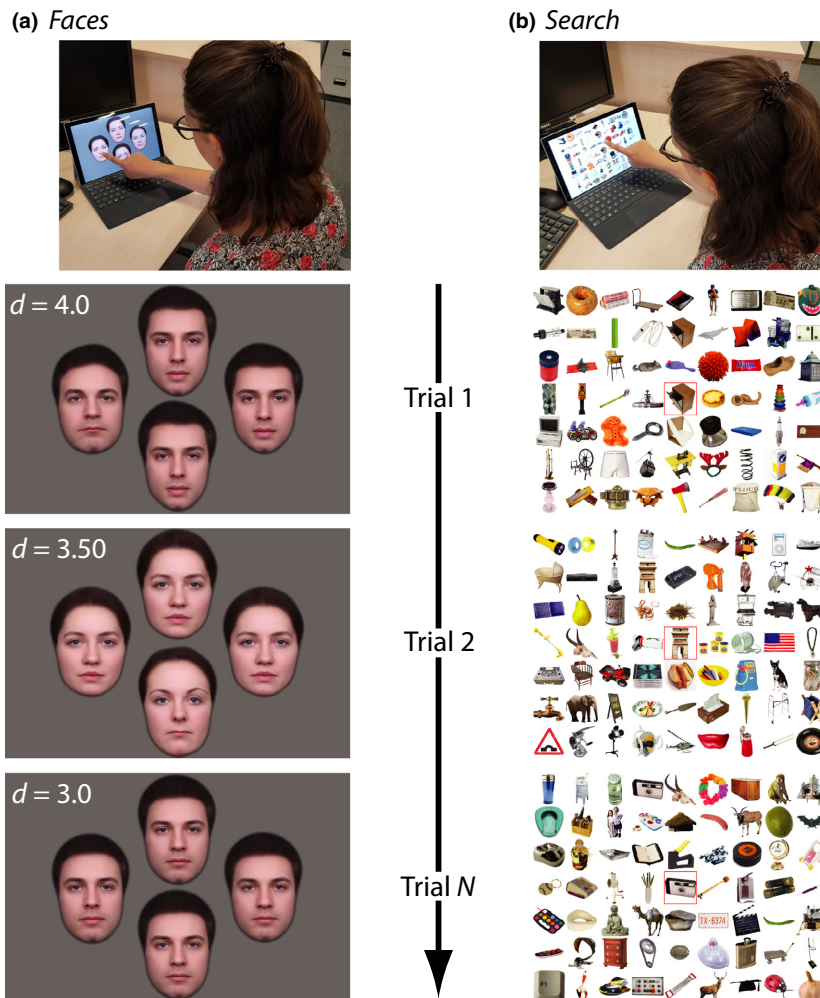
participants prior to testing, and participants were given £15 compensation for their time.

**Equipment**

Both tests were run on a Microsoft Surface Pro 4 ([www.microsoft.com](http://www.microsoft.com)): a touchscreen tablet computer with an IPS (in-plane switching) screen measuring 26 cm × 17.3 cm (Figure 1). Participants were positioned approximately 50 cm from the screen at the start of each test (using a tape measure); however, viewing distance was not strictly constrained. On both tests, the screen was viewed binocularly, and participants were allowed to move their eyes freely. Participants responded by touching the tablet screen, as detailed separately for each test, below.

**Task 1: Faces**

On each trial, the participant was asked to identify (touch) the ‘odd one out’, from a set of four faces (four-alternative forced-choice discrimination; 4AFC). Three of the faces were identical within a given trial (the Standard), though varied randomly between trials. The fourth face (the Target) varied from the Standard by parametrically manipulating photographs of real faces. The source of our face photographs was a dataset from the University of Stirling [Stirling ESRC 3D Face Database: (<http://pics.stir.ac.uk>)] containing 50 frontal images faces [25 male], each manually labelled using Psychomorph software (<http://pics.psych.stir.ac.uk/ESRC/software.html>)<sup>33</sup> with 127 feature points (points on the face that define head shape, hairline and



**Figure 1.** Methods. Setup and example trials for the (a) *Faces* test, and (b) *Search* test. The ‘*d*’ value for the *Faces* test indicate the magnitude of dissimilarity between the target face and the standard faces, with lower values implying more difficult discriminations (see body text for details). In the *Search* test the cue always appeared centrally, within a red box, while the location of the matching target was varied systematically between trials, in random order.

internal feature locations). These 127 features could then be varied by a percentage to create progressively more dissimilar versions of a particular reference image. Faces were masked at presentation to display only the face and removing the ears (which were often observed to exhibit image artefacts after warping). For further technical details regarding the generation and manipulation of the face stimuli see previous works by Logan and colleagues,<sup>34</sup> after which the present work was based.

The Standard face and the location of the Target varied randomly between trials (see *Figure 1a*). There was no time limit, but participants were encouraged to respond “as quickly and accurately as you can”. Depending on whether the participant answered correctly or incorrectly on the previous trial, the degree of similarity between the faces was decreased or increased after every trial, using a QUEST adaptive algorithm.<sup>35</sup> The test ran for a fixed number of 50 trials. However, it was anticipated that fewer trials would be sufficient to obtain a stable estimate of performance (see *Results*).

The outcome measure was *threshold*: the smallest difference between the faces that the participant could detect reliably. This was defined mathematically as the Maximum Likelihood Estimate (MLE) of the mean of the posterior density function, which was computed after every trial. As illustrated in *Figure 1a*, threshold values typically ranged between 3–4, with bigger numbers indicating poorer performance (lower sensitivity). Note that Threshold was measured in units of dissimilarity,  $d$ , which express the Euclidean distance between the Standard and the just-noticeably-different Target face.

## Task 2: Search

On each trial, a random image (the Reference) was presented in the centre of the screen, and remained visible throughout the trial. After 1 s, 62 additional images then appeared also on the screen, spaced uniformly on a  $7 \times 9$  grid (see *Figure 1b*). One of these 62 images (the Target) was identical to the central Reference image. Participants were asked to locate the matching Target image as quickly as possible, and to touch it. All 63 images (including the Reference image) varied on every trial, drawn randomly from a previously described database of 2400 real world objects<sup>36</sup> (see *Figure 1b* for examples).

Unlike with the *Faces* test, the difficulty of the task did not vary trial-by-trial. However, the location of the Target was systematically manipulated: appearing once at each possible grid location over the course of 62 trials. In principle, participants could therefore use a process of elimination from previous trials to inform where the Target was most likely to appear next. In practice, however, there was no evidence that participants attempted

to do so, or were even aware of this possibility. Moreover, the order of target locations was randomised each time the test was run.

Every possible location was tested once (62 trials). However, it was anticipated/hypothesised that a subset of test locations might be sufficient (see *Results*). The outcome measure was **Response Time** (RT), in seconds. This was measured once at each test location, and the overall median RT was computed as a summary measure of performance.

This test represents a modified version of a previous test that we developed to explore the effects of non-neovascular (dry) age-related macular degeneration (AMD).<sup>37</sup> It was modified in the present work to ensure that the Reference image remained visible in the centre of the screen throughout, in order to minimise the memory component of the test. We have in the past also used cluttered, real-world scenes to assess object search performance<sup>18,23</sup> (e.g., using 2D photographs or 3D virtual environments). A uniform grid of discrete objects was employed in the present test, however, as it allowed the target location to be systematically varied between trials, and would allow us to more easily manipulate the stimuli in future (e.g., to refine the stimulus set, or in order to examine a particular domain of object categories).

## Measure of cognition

Participants also completed the “Digit Span” subtest from the Wechsler Adult Intelligence Scale IV (WAIS-IV) test battery. The Digit Span test measures participants’ ability to repeat increasing sequences of numbers forwards and backwards.<sup>38</sup> Low scores may indicate problems with working memory (or a general lack of motivation<sup>39</sup>).

## Analysis

Standard inferential statistical analyses were performed using MathWorks MATLAB R2016b (<https://www.mathworks.com/products/matlab>). When reporting key statistics, bootstrapping was used to compute 95% Confidence Intervals (CI<sub>95</sub>;  $N = 20\,000$ ; bias-corrected and accelerated method).

## Availability of test materials and study data

Study data for both tests are available as Supplemental Material. This dataset also includes the measurements of vision (Acuity, Contrast Sensitivity), cognition, and usability, for each participant.

The *Face* test was programmed in Python ([www.python.org](http://www.python.org)) by author IW, using the OpenSesame toolbox<sup>40</sup> ([www.osdoc.cogsci.nl](http://www.osdoc.cogsci.nl)). The source code is freely available

online for non-commercial use at: <https://www.bitbucket.org/iainwilson/facediscrimination>.

The *Search* test was programmed in C# by author WB, and an executable is freely available online for non-commercial use at: <https://github.com/CrabbLab/CrazySearch>.

### Results

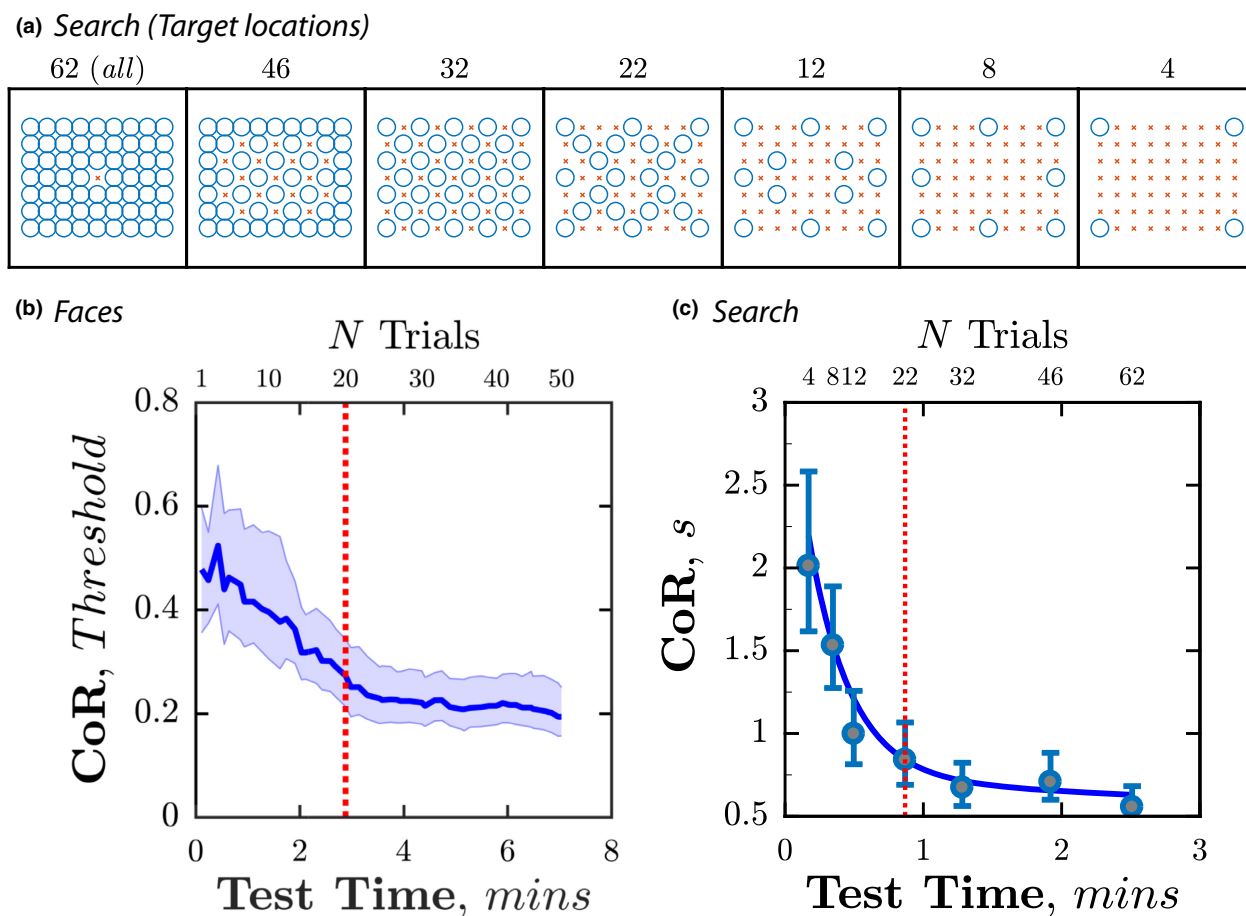
#### Test refinement

Each test was intentionally run for longer than piloting indicated was necessary. To determine how many trials were actually required to obtain stable estimates of performance, we examined how test-retest variability (the 95% Coefficient of Repeatability; CoR) varied as a function of test duration/number of trials (*Figure 2*). For *Faces*, this involved simply analysing the first  $N$  trials (i.e., since the adaptive algorithm provides an updated, maximum-likelihood estimate of discrimination ability after every trial). For *Search*, data were analysed from progressively more

sparse subsets of spatially-distributed locations, as shown in *Figure 2a*.

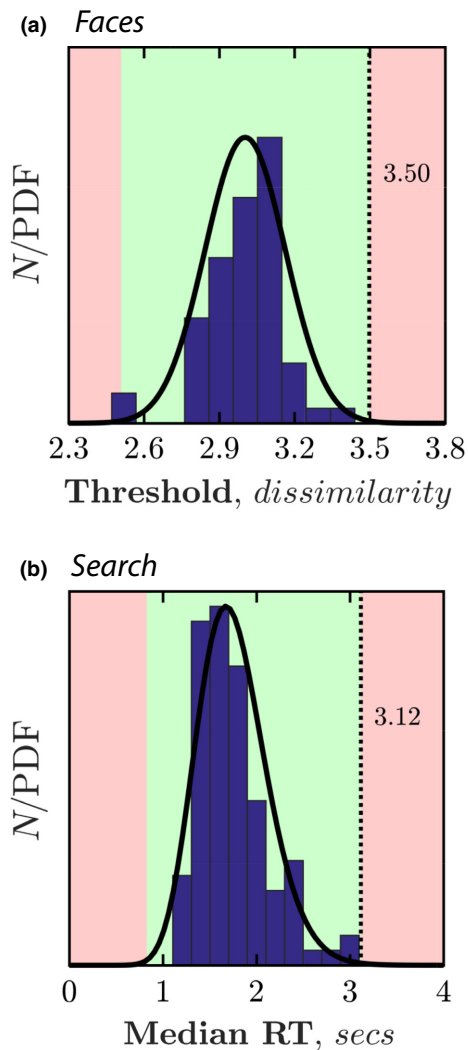
Unsurprisingly, increasing the number of trials resulted in greater measurement precision (*Figure 2b-c*). However, precision improved rapidly for the first 20 (*Faces*) or 22 trials (*Search*), and more gradually thereafter. For example, in the *Faces* task CoR<sub>95</sub> decreased by 43% as the number of trials increased from  $N = 1$  to  $N = 20$ , with only a further 17% reduction by  $N = 50$ . Similarly, in the *Search* task CoR<sub>95</sub> decreased by 58% from  $N = 4$  to  $N = 22$ , with a further 14% reduction by  $N = 62$ .

The ideal test duration will depend on the level of precision required. However, we anticipate that ~20 trials will be sufficient for most clinical purposes. This corresponds to approximately 1–3 min (see below). Tests of a longer duration would also likely be unacceptable to people waiting in routine clinics. We therefore report data only for these subsets of ~20 trials in the remainder of the manuscript.



**Figure 2.** Test refinement. Within-subject (test-retest) measurement variability, as a function of  $N$  trials for the (b) *Faces* test and (c) *Search* Test. Coefficients of Repeatability were derived using Bland-Altman analysis, as detailed in Figure 5, below. Panel (a) shows the target locations (blue circles) associated with each *Search* grid.





**Figure 3.** Normative data for the (a) *Faces* test, and (b) *Search* test. Curves show maximum likelihood fits of appropriate probability density functions (*Faces*: Gaussian PDF. *Search*: Gamma PDF). Dashed vertical lines indicate the cutoff point, below which 99% of normally-sighted participants would be expected to score.

### Normative values

There was no systematic difference in performance between the first and second run, either for *Faces* ( $p = 0.11$ ) or *Search* ( $p = 0.42$ ). Accordingly, data from both runs were concatenated to produce the normative distributions shown in *Figure 3*. Appropriate probability distributions (black lines) were fitted to the raw data. These were used to determine the 99% upper-bound point (dashed vertical line): the cut-off point below which 99% of young, visually-normal participants would be expected to score. These values were 3.50 (*Faces*) and 3.1 s (*Search*). Values greater than this may indicate abnormal test performance.

For *Search* (only) it is also possible to consider performance for individual spatial locations. This could be important if, for example, attempting to detect localised visual field loss. Accordingly, *Figure 4* shows normative median values, and the 99% upper cut-off value for each location (computed in the same way as for the overall median RT, in *Figure 3b*). As highlighted in *Figure 4b*, there was a clear effect of eccentricity, with participants being slower to locate more peripheral targets.

### Test-retest reliability

As shown in *Figure 5*, the Coefficient of Repeatability  $\{\pm \text{CI } 95\%$  was 0.27  $\{0.22, 0.35\}$  for *Faces*, and 0.84  $\{0.71, 1.07\}$  for *Search*. As is evident by inspection, there was no systematic effects of learning or fatigue. Measurement error tended to be approximately normally distributed, although on the *Search* task there was a tendency for variability to increase as a function of overall reaction time.

### Test duration

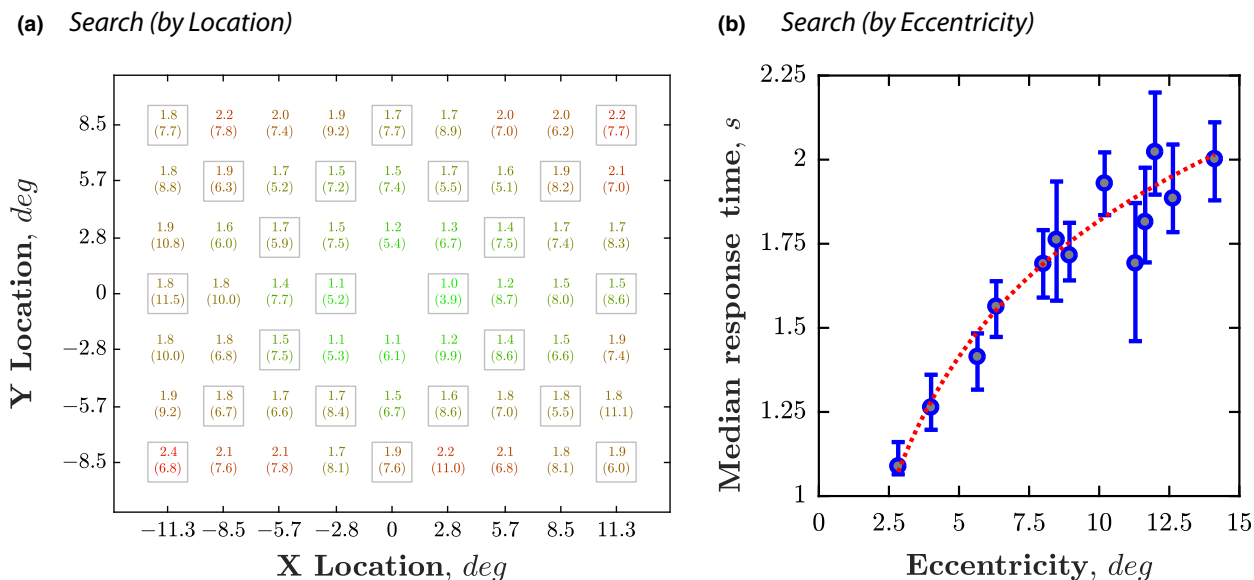
For *Faces*, median  $\{\text{CI}_{95\%}\}$  test duration was 191  $\{168, 228\}$  s for the first run, and 155  $\{139, 186\}$  for the second: a statistically significant difference ( $t$ -test:  $p = 0.042$ ) of 19%. For *Search*, median  $\{\text{CI}_{95\%}\}$  test duration was 51  $\{46, 56\}$  s for the first run, and 47  $\{45, 50\}$  s for the second: a non-significant difference ( $t$ -test:  $p = 0.16$ ).

### Usability & completion rate

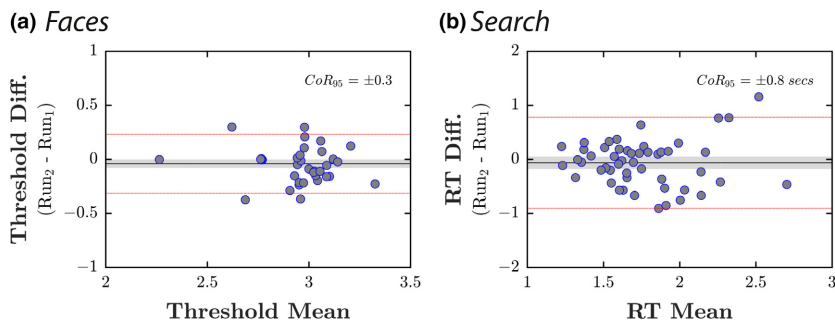
All participants (100%) completed both tests twice, with no participants exhibiting/reporting any difficulties. Participants were asked to rate how clearly they understood what to do on each test, on a scale from 0 (incomprehensible) to 10 (very understandable). Ratings of comprehension were remarkably high for both tests with a mean  $\{\text{CI}_{95\%}\}$  rating of 9.6  $\{9.1, 9.8\}$  for *Faces*, and 9.7  $\{9.2, 9.8\}$  for *Search*.

### Relationships with cognition and basic vision

As shown in *Figure 6*, there was no significant association between performance on either test and with: (1) Digit Span general cognition; (2) logMar letter acuity; or Pelli-Robson contrast sensitivity (see *Figure 6* for  $p$  values) – although there was a trend towards an association between acuity and performance on the *Search* task ( $p = 0.056$ ;  $r = 0.27$ ). In those 30 participants who performed both tests, there was also no correlation between scores on the *Faces* and *Search* task ( $r = 0.18$ ;  $p = 0.16$ ).



**Figure 4.** Pointwise normative data for the Search task, as a function of (a) Screen location; (b) Eccentricity from the centre. Grey boxes indicate the subset of 22 points that were used in all other figures and analyses. Data for the other 40 locations also given for completeness. Eccentricities computed assuming a viewing distance of 50 cm. Red dashed line indicates the best fitting power function [ $y = -13.14x^{-0.05} + 13.56$ ; Adjusted  $R^2 = 0.90$ ]. Individual markers indicate median response times with 95% confidence intervals.



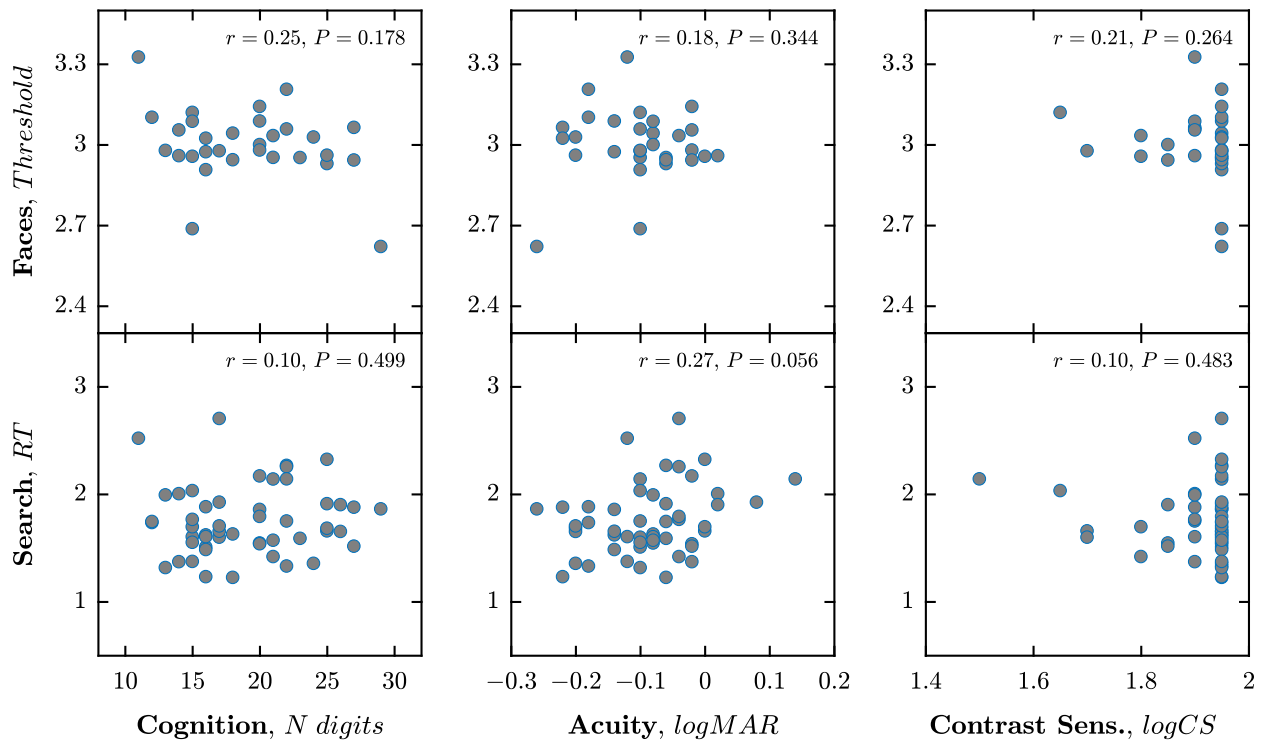
**Figure 5.** Bland-Altman analyses of retest repeatability for (a) Faces and (b) Search. Each marker represents a single participant. Grey shaded regions show 95% confidence intervals around the mean. Dashed red lines indicate the 95% limits of agreement. Text (top-right) gives the 95% Coefficient of Repeatability ( $CoR_{95}$ ).

**Discussion**

The purpose of the present study was to describe, refine, evaluate (in normally sighted young adults), and provide normative control data for two tablet-based tests of real-world visual function, both of which we have made freely available online. The results established that both tests were capable of providing stable estimates of visual function in around 20 trials (~1–2 min). They also defined cutoff points for what constitutes ‘normal’ performance, and showed that the tests could be performed easily by young adults, who, anecdotally, often regarded them as ‘games’ (in contrast to the vision screening ‘tests’ that preceded them). This last point is crucial, as tests must be

simple, intuitive, and engaging if to be completed autonomously by patients, without the need for costly technicians.<sup>41</sup>

There was no relationship between performance on the tablet tests, and scores on a test of cognitive function (*Digit Span*). This is encouraging, as it suggests that they are measuring genuine perceptual abilities, rather than general cognitive ability or effort. However, this result should be taken with caution, since no participants scored outside of normal limits on the cognitive test, and all behavioural measures are inevitably susceptible, to some degree, to cognitive factors such as motivation, understanding, and compliance. Likewise, there was little or no association with more basic tests of vision, suggesting that these novel



**Figure 6.** Scatter plots showing the relationships with cognition (Digital Recall) and basic vision (Acuity, Contrast Sensitivity). Each marker indicates a single participant, with scores across the two runs of each test mean-averaged. Text (top-right) gives the results of independent Spearman's rank correlations.

measures may be able to provide additional information over and above what can be inferred from standard measures of acuity or contrast sensitivity. Again though, this may in part reflect the highly homogenous nature of the cohort; in the limiting case it is, for example, necessary to have some basic level of acuity to be able to even attempt to perform either test. Ultimately, additional data from patients are required to establish the clinical utility of these measures, over and above more basic measures of visual function (see *Limitations*, below).

### Potential applications

The new tablet-based tests described here are simple and relatively enjoyable, do not require a trained operator to supervise, and can be run using inexpensive commercial equipment that is easy to maintain or replace. They would therefore be ideal for giving to patients in waiting rooms, where many individuals would likely welcome the distraction, and where the tests can effectively provide 'free' data – without extending appointment durations or further burdening clinical staff. For example, we are currently examining whether they can be used in diabetic macular edema [DME] clinics to identify individuals experiencing real-world difficulties despite mild VA loss.<sup>42</sup> We are also

exploring their use with age-related macular degeneration (AMD) patients to assess the real-world impact of progressive central vision loss. Ultimately, however, we envisage their potential applications to be multifarious. As such, we have made the tests freely available online for people to use and develop (see *Methods*).

### Comparison to previous literature

The present work is timely, as a number of other tablet-based tests have recently been reported, designed to measure various more basic aspects of visual function, such as visual acuity,<sup>15,41,43,44</sup> contrast sensitivity,<sup>45–49</sup> visual fields,<sup>12,50–56</sup> stereopsis,<sup>57</sup> and colour vision.<sup>58,59</sup> What distinguishes the tests described in the present study is that they are intended to measure 'high-level' function on everyday tasks. They can therefore be thought of as a complementary, more objective analog to traditional PROMs, and so may be particularly well suited to flagging up those patients who are experiencing everyday difficulties not captured by more basic measures of visual function, such as acuity or contrast sensitivity.<sup>42,60</sup>

On a practical level, the novel measures described in the present work are also much easier to administer than many of these more basic measures, since we anticipate there will



be no need to calibrate the luminance or chromaticity or display (i.e., unlike contrast sensitivity), or to precisely control the viewing distance or ambient lighting of the observer (i.e., unlike acuity). Though it is perhaps interesting to note that both low- and high-level functional tests could be potentially be performed with the same tablet devices. For example, we describe elsewhere a tablet-perimeter (Eye-catcher) that uses the exact same tablet computer as described in the present work.<sup>56</sup> Elsewhere, the use of tablets as a means of sharing clinical data between a patient's eyecare team has also been explored.<sup>61</sup>

With respect to traditional PROMs, it is worth stressing that the present tests are not intended as a like-for-like replacement. Asking patients directly remains the best way to ascertain how someone feels about their condition, and PROMs, when used well,<sup>62</sup> have been shown capable of providing important insights into the everyday difficulties that patients face.<sup>63,64</sup> PROMs are, however, potentially limited by individual differences in personality, knowledge and expectations of disease,<sup>65</sup> as well as by differences in lifestyle (e.g., with some individuals not reporting difficulties with a particular task domain because they now avoid it altogether, or have develop adaptation strategies to cope with their condition). We might therefore learn something complementary by assessing actual performance in surrogates of tasks that patients would encounter every day.<sup>66</sup>

### Limitations & future work

The primary limitation of the present study is that we only assessed young people with healthy vision. These data allowed us to refine the measures, perform a preliminary assessment of feasibility, and provide limits on what constitutes normal performance. In future, however, it will be necessary to collect data from patients to more fully assess their speed, reliability, and relationship to basic measures of visual function. Doing so may also suggest further refinements. For example, the *Faces* test, with a median test duration of around 3 min may already be 'at the limit' of what is clinically practicable, and it may be necessary to further shorten the test, either by reducing the number of trials, and/or by integrating a Bayesian prior into the adaptive algorithm.<sup>67</sup>

A related limitation is that we only examined young adults. Older adults are often slower and less accurate at locating objects in cluttered scenes.<sup>68–70</sup> In future, it would therefore be helpful to collect additional normative data for older adults performing the *Search* task. There may also be age-related changes in performance on the *Faces* test also. Although interestingly older adults, while sometimes exhibiting difficulties recalling faces,<sup>71</sup> often appear no worse than younger adults at discriminating between

simultaneously presented faces, after correcting for differences in vision.<sup>72</sup>

A final limitation is that testing in the present work took place in a controlled, university environment. This is potentially quite different to a busy clinic. In future, it will be necessary to conduct more extensive studies to assess the feasibility — and clinical utility — of deploying such tests in a practicing clinic. These are questions which we currently investigating, and intend to report data from in due course.

### Acknowledgments

This study was funded as part of an unrestricted investigator-initiated research grant from Allergan. The funding organisation had no role in the design or conduct of this research.

### Conflict of interest

The authors report the following professional relationships, and no conflicts of interest: PR Jones, None; I Tigchelaar, None; G Demaria, None; I Wilson, None; W Bi, None; DJ Taylor, None; DP Crabb, Roche, Allergan (F), Allergan, Santen, THEA, Bayer(R).

### References

1. Denniston AK, Kyte D, Calvert M & Burr JM. An introduction to patient-reported outcome measures in ophthalmic research. *Eye* 2014; 28: 637–645.
2. Azuara-Blanco A, Burr J, Ramsay C *et al.* Effectiveness of early lens extraction for the treatment of primary angle-closure glaucoma (EAGLE): a randomised controlled trial. *Lancet* 2016; 388: 1389–1397.
3. Chakravarthy U, Harding SP, Rogers CA *et al.* Ranibizumab versus bevacizumab to treat neovascular age-related macular degeneration: one-year findings from the IVAN randomized trial. *Ophthalmology* 2012; 119: 1399–1411.
4. Krezel AK, Hogg RE & Azuara-Blanco A. Patient-reported outcomes in randomised controlled trials on age-related macular degeneration. *Br J Ophthalmol* 2015; 99: 1560–1564.
5. Gazzard G, Konstantakopoulou E, Garway-Heath D *et al.* Selective laser trabeculoplasty versus eye drops for first-line treatment of ocular hypertension and glaucoma (LiGHT): a multicentre randomised controlled trial. *Lancet* 2019; 393: 1505–1516.
6. Friedman SM, Munoz B, Rubin GS, West SK, Bandeen-Roche K & Fried LP. Characteristics of discrepancies between self-reported visual function and measured reading speed. Salisbury Eye Evaluation Project Team. *Invest Ophthalmol Vis Sci* 1999; 40: 858–864.

7. McGwin G Jr, Owsley C & Ball K. Identifying crash involvement among older drivers: agreement between self-report and state records. *Accid Anal Prev* 1998; 30: 781–791.
8. Hochberg C, Maul E, Chan ES *et al.* Association of vision loss in glaucoma and age-related macular degeneration with IADL disability. *Invest Ophthalmol Vis Sci* 2012; 53: 3201–3206.
9. Rovner BW & Casten RJ. Neuroticism predicts depression and disability in age-related macular degeneration. *J Am Geriatr Soc* 2001; 49: 1097–1100.
10. Rovner BW, Casten RJ, Hegel MT *et al.* Personality and functional vision in older adults with age-related macular degeneration. *J Vis Impair Blind* 2014; 108: 187–199.
11. Dabasia PL, Edgar DF, Garway-Heath DF & Lawrenson JG. A survey of current and anticipated use of standard and specialist equipment by UK optometrists. *Ophthalmic Physiol Opt* 2014; 34: 592–613.
12. Nesaratnam N, Thomas PBM, Kirolos R, Vingrys AJ, Kong GYX & Martin KR. Tablets at the bedside-iPad-based visual field test used in the diagnosis of Intrasellar Haemangiopericytoma: a case report. *BMC Ophthalmol* 2017; 17: 53.
13. Glen FC, Baker H & Crabb DP. A qualitative investigation into patients' views on visual field testing for glaucoma monitoring. *BMJ Open* 2014; 4: e003996.
14. Strutton R, Du Chemin A, Stratton IM & Forster AS. System-level and patient-level explanations for non-attendance at diabetic retinopathy screening in Sutton and Merton (London, UK): a qualitative analysis of a service evaluation. *BMJ Open* 2016; 6: e010952.
15. Bastawrous A, Rono HK, Livingstone IAT *et al.* Development and validation of a smartphone-based visual acuity test (peek acuity) for clinical practice and community-based fieldwork. *JAMA Ophthalmol* 2015; 133: 930–937.
16. Glen F, Smith ND & Crabb DP. Saccadic eye movements and face recognition performance in patients with central glaucomatous visual field defects. *Vision Res* 2013; 82: 42–51.
17. Glen FC, Crabb DP, Smith ND, Burton R & Garway-Heath DF. Do patients with glaucoma have difficulty recognizing faces? *Invest Ophthalmol Vis Sci* 2012; 53: 3629–3637.
18. Smith ND, Crabb DP & Garway-Heath DF. An exploratory study of visual search performance in glaucoma. *Ophthalmic Physiol Opt* 2011; 31: 225–232.
19. Smith ND, Glen FC & Crabb DP. Eye movements during visual search in patients with glaucoma. *BMC Ophthalmol* 2012; 12: 45.
20. Mazzoli LS, Urata CN & Kasahara N. Face memory deficits in subjects with eye diseases: a comparative analysis between glaucoma and age-related macular degeneration patients from a developing country. *Graefe's Arch Clin Exp Ophthalmol* 2019; 257: 1941–1946.
21. Roux-Sibilon A, Rutgé F, Aptel F *et al.* Scene and human face recognition in the central vision of patients with glaucoma. *PLoS ONE* 2018; 13: e0193465.
22. Taylor DJ, Smith ND, Binns AM & Crabb DP. The effect of non-neovascular age-related macular degeneration on face recognition performance. *Graefe's Arch Clin Exp Ophthalmol* 2018; 256: 815–821.
23. Taylor DJ, Smith ND & Crabb DP. Searching for objects in everyday scenes: measuring performance in people with dry age-related macular degeneration. *Invest Ophthalmol Vis Sci* 2017; 58: 1887–1892.
24. Crabb DP & Taylor DJ. Searching for unity: real-world versus item-based visual search in age-related eye disease. *Behav Brain Sci* 2017; 40: e135.
25. Mulcahey MJ, Haley SM, Duffy T, Pengsheng N & Betz RR. Measuring physical functioning in children with spinal impairments with computerized adaptive testing. *J Pediatr Orthop* 2008; 28: 330–335.
26. Payne M, Janzen S, Earl E, Deathe B & Viana R. Feasibility testing of smart tablet questionnaires compared to paper questionnaires in an amputee rehabilitation clinic. *Prosthet Orthot Int* 2017; 41: 420–425.
27. Patel V, Hale TM, Palakodeti S, Kvedar JC & Jethwani K. Prescription tablets in the digital age: a cross-sectional study exploring patient and physician attitudes toward the use of tablets for clinic-based personalized health care information exchange. *JMIR Res Protoc* 2015; 4: e116.
28. Stribling JC & Richardson JE. Placing wireless tablets in clinical settings for patient education. *J Med Libr Assoc JMLA* 2016; 104: 159.
29. Brinker TJ, Brieske CM, Esser S *et al.* A face-aging app for smoking cessation in a waiting room setting: pilot study in an HIV outpatient clinic. *J Med Internet Res* 2018; 20: e10976.
30. Hassan R, Twyman NW, Nah FFH & Siau K. Patient engagement in the medical facility waiting room using Gamified healthcare information delivery. In: *HCI in Business, Government, and Organizations: Information Systems*, Volume 9752. Cham: Springer, 2016, (Nah FH & Tan CH, eds), pp. 412–423.
31. Kelly EA, Stadler ME, Nelson S, Runge CL & Friedland DR. Tablet-based screening for hearing loss: feasibility of testing in nonspecialty locations. *Otol Neurotol* 2018; 39: 410–416.
32. Elliott DB, Bullimore MA & Bailey IL. Improving the reliability of the Pelli-Robson contrast sensitivity test. *Clin Vision Sci* 1991; 6: 471–475.
33. Tiddeman BP, Stirrat MR & Perrett DI. Towards realism in facial image transformation: results of a wavelet MRF method. *Comput Graph Forum* 2005; 24: 449–456.
34. Logan AJ, Wilkinson F, Wilson HR, Gordon GE & Loffler G. The Caledonian face test: a new test of face discrimination. *Vision Res* 2016; 119: 29–41.
35. Watson AB & Pelli DG. QUEST: a Bayesian adaptive psychometric method. *Percept Psychophys* 1983; 33: 113–120.
36. Brady TF, Konkle T, Alvarez GA & Oliva A. Visual long-term memory has a massive storage capacity for object details. *Proc Natl Acad Sci* 2008; 105: 14325–14329.

37. Higgins BE, Taylor DJ, Bi W, Binns AM & Crabb DP. Novel computer-based tests for assessing performance in visually guided tasks in people with age-related macular degeneration: searching for everyday objects and detecting road signs. *Invest Ophthalmol Vis Sci* 2019; 60: 5922.
38. Banken JA. Clinical utility of considering digits forward and digits backward as separate components of the wechsler adult intelligence scale-revised. *J Clin Psychol* 1985; 41: 686–691.
39. Axelrod BN, Fichtenberg NL, Millis SR & Wertheimer JC. Detecting incomplete effort with digit span from the Wechsler Adult Intelligence Scale – Third Edition. *Clin Neuropsychol* 2006; 20: 513–523.
40. Mathôt S, Schreij D & Theeuwes J. OpenSesame: an open-source, graphical experiment builder for the social sciences. *Behav Res Methods* 2012; 44: 314–324.
41. Aslam TM, Parr NRA, Murray IJ *et al.* Development and testing of an automated computer tablet-based method for self-testing of high and low contrast near visual acuity in ophthalmic patients. *Graefe's Arch Clin Exp Ophthalmol* 2016; 254: 891–899.
42. Gonder JR, Walker VM, Barbeau M *et al.* Costs and quality of life in diabetic macular edema: Canadian Burden of Diabetic Macular Edema Observational Study (C-REALITY). *J Ophthalmol* 2014; 2014: 207–215.
43. Black JM, Jacobs RJ, Phillips G *et al.* An assessment of the iPad as a testing platform for distance visual acuity in adults. *BMJ Open* 2013; 3: e002730.
44. Zhang Z, Zhang S, Huang X & Liang L. A pilot trial of the iPad tablet computer as a portable device for visual acuity testing. *J Telemed Telecare* 2013; 19: 55–59.
45. Wu Z, Guymer RH, Jung CJ *et al.* Measurement of retinal sensitivity on tablet devices in age-related macular degeneration. *Transl Vis Sci Technol* 2015; 4: 13.
46. Dorr M, Lesmes LA, Elze T, Wang H, Lu Z-L & Bex PJ. Evaluation of the precision of contrast sensitivity function assessment on a tablet device. *Sci Rep* 2017; 7: 46706.
47. Dorr M, Lesmes LA, Lu ZL & Bex PJ. Rapid and reliable assessment of the contrast sensitivity function on an iPad. *Invest Ophthalmol Vis Sci* 2013; 54: 7266–7273.
48. Bodduluri L, Boon MY, Ryan M & Dain SJ. Normative values for a tablet computer-based application to assess chromatic contrast sensitivity. *Behav Res Methods* 2018; 50: 673–683.
49. Rodriguez-Vallejo M, Remón L, Monsoriu JA & Furlan WD. Designing a new test for contrast sensitivity function measurement with iPad. *J Optom* 2015; 8: 101–108.
50. Anderson AJ, Bedgood PA, Kong YXG, Martin KR & Vingrys AJ. Can home monitoring allow earlier detection of rapid visual field progression in glaucoma? *Ophthalmology* 2017; 124: 1735–1742.
51. Schulz AM, Graham EC, You Y, Klistorner A & Graham SL. Performance of iPad based threshold perimetry in glaucoma and controls. *Clin Exp Ophthalmol* 2018; 46: 346–355.
52. Kong YXG, He M, Crowston JG & Vingrys AJ. A comparison of perimetric results from a tablet perimeter and Humphrey field analyzer in glaucoma patients. *Transl Vis Sci Technol* 2016; 5: 2.
53. Vingrys AJ, Healey JK, Liew S *et al.* Validation of a tablet as a tangent perimeter. *Transl Vis Sci Technol* 2016; 5: 3.
54. Prea SM, Kong YXG, Mehta A *et al.* Six-month longitudinal comparison of a portable tablet perimeter with the Humphrey Field Analyzer. *Am J Ophthalmol* 2018; 190: 9–16.
55. Johnson CA, Thapa S, Kong YXG & Robin AL. Performance of an iPad application to detect moderate and advanced visual field loss in Nepal. *Am J Ophthalmol* 2017; 182: 147–154.
56. Jones PR, Smith ND, Bi W & Crabb DP. Portable perimetry using eye-tracking on a tablet computer – a feasibility assessment. *Transl Vis Sci Technol* 2019; 8: 17.
57. Rodriguez-Vallejo M, Ferrando V, Montagud D, Monsoriu JA & Furlan WD. Stereopsis assessment at multiple distances with an iPad application. *Displays* 2017; 50: 35–40.
58. de Fez D, Luque MJ, Matea L, Piñero DP & Camps VJ. New iPad-based test for the detection of color vision deficiencies. *Graefe's Arch Clin Exp Ophthalmol* 2018; 256: 2349–2360.
59. Alvaro L, Alvarez J, Meyer B *et al.* Tablet-based app for screening for colour vision deficiencies in young children. *Perception* 2019; 48: 46.
60. Drum B, Calogero D & Rorer E. Assessment of visual performance in the evaluation of new medical products. *Drug Discov Today Technol* 2007; 4: 55–61.
61. Kern C, Fu DJ, Kortuem K *et al.* Implementation of a cloud-based referral platform in ophthalmology: making telemedicine services a reality in eye care. *Br J Ophthalmol* 2019. <https://doi.org/10.1136/bjophthalmol-2019-314161>. [Epub ahead of print]
62. Skalicky SE, Lamoureux EL, Crabb DP & Ramulu PY. Patient-reported outcomes, functional assessment, and utility values in glaucoma. *J Glaucoma* 2019; 28: 89–96.
63. Ramulu P. Glaucoma and disability: which tasks are affected, and at what stage of disease? *Curr Opin Ophthalmol* 2009; 20: 92–98.
64. Medeiros FA, Gracitelli CPB, Boer ER, Weinreb RN, Zangwill LM & Rosen PN. Longitudinal changes in quality of life and rates of progressive visual field loss in glaucoma patients. *Ophthalmology* 2015; 122: 293–301.
65. Somner JEA, Sii F, Bourne RR, Cross V, Burr JM & Shah P. Moving from PROMs to POEMs for glaucoma care: a qualitative scoping exercise. *Invest Ophthalmol Vis Sci* 2012; 53: 5940–5947.
66. Crabb DP. A view on glaucoma – are we seeing it clearly? *Eye* 2016; 30: 304–313.
67. Watson AB. QUEST+: a general multidimensional Bayesian adaptive psychometric method. *J Vis* 2017; 17: 10.

68. Ho G, Scialfa CT, Caird JK & Graw T. Visual search for traffic signs: the effects of clutter, luminance, and aging. *Hum Factors* 2001; 43: 194–207.
69. Grahame M, Laberge J & Scialfa CT. Age differences in search of Web pages: the effects of link size, link number, and clutter. *Hum Factors* 2004; 46: 385–398.
70. Brennan AA, Bruderer AJ, Liu-Ambrose T, Handy TC & Enns JT. Lifespan changes in attention revisited: everyday visual search. *Can J Exp Psychol Can Psychol expérimentale* 2017; 71: 160–171.
71. Boutet I & Faubert J. Recognition of faces and complex objects in younger and older adults. *Mem Cognit* 2006; 34: 854–864.
72. Barnes CS, De l'Aune W & Schuchard RA. A test of face discrimination ability in aging and vision loss. *Optom Vis Sci* 2011; 88: 188–199.

### Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Data S1.** Raw data for each participant, including primary outcome measures for the faces and search task, as well as measurements of vision (Visual Acuity, Contrast Sensitivity), cognition, and usability.