Measuring Low Vision: Perception and Function

A Dissertation
SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA
BY

Andrew C. Freedman

IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Advisor: Gordon Legge

May 2024

Acknowledgement

During my doctoral studies, I've learned that while one can imagine all the research they like on their own, a strong support network is a vital part of actually getting the science done. I have a great many people to thank for helping to bring this dissertation to fruition.

First and foremost, I offer Professor Gordon E. Legge my sincerest thanks for his central role throughout my Ph. D. studies. Gordon has been an exceptional advisor, demonstrating how to conduct science with a strong sense of purpose, intellectual rigor and honesty, and focused clarity. He has been an invaluable mentor, ensuring I never lost track of the meaning behind our work, offering support and patience when progress stalled, and fostering an environment of real comradery in our lab – a well of stability and companionship in tumultuous times. I could go on, and on, and on, but one of the major lessons Gordon has taught me is the value of brevity in one's writing, so suffice it to say that it has truly been a pleasure to work alongside him.

I would next like to acknowledge my pride in being a member of GELLAB, due in no small part to the company I kept there, including my friends Rachel Gage, Dr. Siyun Liu, Dr. Walter Wu, Dr. Yingzi Xiong, Dr. Nilsu Atilgan, Dr. Yingchen He, Christina Devine, and Dr. Quan Lei. Whether I found you underground in the basement of Elliott Hall, starting an argument at the Purple Onion, or exploring the city surrounding a conference, you were all reliable sources of my best memories from grad school. In particular, I'd like to thank Rachel for acting as the backbone of the work done for this dissertation (again, brevity, but you know the many things that you did), and Siyun for guiding me through the maze of programming in VR.

I am also grateful to my thesis committee members, Professors Dan Kersten, Peggy Nelson, and Stephen Engel for their roles in both my preliminary and final oral examinations. Knowing that I had the opportunity to share my research with a group of keenly insightful scientists positioned my defense as something to look forward to, rather than as simply another hurdle.

To everyone who participated in the experiments included in this dissertation, you have my thanks as well. Without your effort, these studies would quite literally have been impossible.

Support came from outside the world of academia too – to my many friends who got to (had to?) watch this journey from the outside, thank you for sticking with me while I figured out how being a scientist works. It has taken some time, but it looks like we've all been able to find our paths.

Perhaps no other group provided more support than my family. Unwavering in your assuredness that I would make it through, even in the most trying of times, I thank you all for your support. Mom, it looks like I really did manage to fulfil the promise I made in your mother's day card years ago – thanks for always believing I would. Dad, I know you

would be proud to read this – I'm sure everyone you knew would have heard about it, likely several times. To Alyssa, Teighan, Dallas, and Brandon, I want to thank you for always having open ears and homes open, and for playing games when there was time for it. Finally, I send my utmost gratitude to my delightful wife, Dr. Yihan Wu. Thank you for your stalwart love and support when the work was hardest, for your keen insights when mine ran dry, and for helping me to never forget how important this work is.

Lastly, I acknowledge that my work is intrinsically dependent on the countless insights from other scientists who came before me. My designs and analyses rest on the foundation of their work, standing on the shoulders of giants. My wish is that this work reinforces that bedrock, and serves as a launch point for successive scientists who share my goal of utilizing science to improve the wellbeing of all people.

Abstract

The impact of vision impairment affects both what can be perceived as well as how effectively visual tasks can be carried out. Therefore, a complete understanding of the effects of vision impairment must account for both its functional and sensory impact. This dissertation investigates both sides, focusing on function in the context of a visual search task, and focusing on sensation in the context of remote tests of visual acuity and contrast sensitivity. Chapter 1 provides an overview of the thesis and introduces key concepts employed throughout the research. Chapter 2 evaluates how visual impairment affects functioning in the context of a visual search task, focusing on how illumination and contrast affect object visibility, a key factor determining whether a space is safe and easy to move through with impaired vision. While recommendations for improving object visibility typically center on increasing illumination, suggesting "brighter is better," we aimed to determine whether the efficacy of such interventions depend on the degree to which they also increase contrast. Normally sighted participants wore blur goggles that severely restricted their visual acuity, artificially restricting it to 20/1250. They then walked through a room, visually searching for target objects, some of which received enhanced illumination from spotlights. Results showed that enhanced illumination improved the rate at which targets were seen only if it also improved the target's contrast with the background behind it, and this information was used to provide guidelines for implementing focused lighting for visual accessibility. Chapter 3 focuses on evaluating the sensory impact of visual impairment by describing the development and initial validation process for novel remote tests of visual acuity and contrast sensitivity. Remote

testing allows patients to undergo vision evaluation from a location outside a clinic, and typically involves the digital presentation of test stimuli over a device the patient owns. Such tests are especially valuable during times when travel is restricted (e.g. COVID-19 lockdown), or for individuals with limited mobility, as the tests can be taken without leaving home. The new tests can measure much poorer acuities than other currently available remote tests, and overcome challenges in contrast calibration not addressed by other digital contrast sensitivity tests. To validate the new tests, normally sighted participants with artificially restricted visual acuity and contrast sensitivity were tested in a laboratory environment using both the remote tests and gold-standard clinical charts for comparison. The novel tests showed good agreement with the charts, demonstrating similar scores with low variance. Chapter 4 reports on the continued validation process of the tests developed in Chapter 3, moving from a controlled laboratory testing environment to remote testing of low vision participants at their homes. We found that the tests demonstrated similar scores on average, but scores were also more variable compared to testing in a laboratory environment, potentially due to uncontrolled viewing conditions in the home environment. Altogether, the studies presented in this dissertation provide insight into how the complex functional and sensory consequences of vision impairment can be both measured and addressed. The guidelines and tools developed in this work serve as valuable resources, applicable to both the creation of visually accessible spaces and the administration of remote vision testing.

Table of Contents

List of Tables

List of Figures

Chapter 1: Overview

For most people, vision is integral to daily life and functioning, and a defining aspect of conscious experience. Consequently, chronic impairment of vision that cannot be corrected with glasses, also known as low vision (LV), affects not only what can be seen, but also the myriad of daily functions that rely on vision. Therefore, a complete understanding of the impact of LV must incorporate both how it affects both what is perceived as well as how that information is used. This dissertation contributes to that understanding by studying both how vision impairment affects one's ability to find objects in their surroundings and by developing and evaluating a tool for quantifying what can be seen with impaired vision.

LV affects many aspects of daily functioning, including increased difficulty with mobility, characterized by an increased risk for injurious falls during self-guided travel (Hegde et al., 2017). This is due in part to difficulty avoiding obstacles, particularly those with low luminance (Kuyk et al., 1996; Chung et al., 2018; Grech et al., 2017) and low contrast with their surroundings (Hassan et al., 2007; Lukman et al., 2020). Luminance is a measure of how much light is reflected off the surface of an object from a given perspective (Thompson et al., 2022), while contrast is a measure of difference in luminance between an object and its surroundings. Together, these factors contribute to how bright an obstacle appears, and how much it stands out from its surroundings.

The most common metric used to describe what can be seen with LV is visual acuity (VA), which describes the resolution of vision, or the ability to distinguish fine

details in a visual scene. VA is frequently measured by asking a viewer to read letters of decreasing size on a letter chart, with acuity being measured as the smallest letters that can be resolved at a specified viewing distance. The chart most typically referenced in popular culture is a Snellen chart, with a single large letter E on the top line, though a more modern chart used clinically and in research is the Early Treatment of Diabetic Retinopathy (ETDRS) chart, both of which are shown in Figure 1-1, Panels A and B. Another modern method frequently used in digital acuity tests is the Tumbling E, illustrated in Figure 1-1, Panel C. In a Tumbling E test, an uppercase letter E is presented in one of four orientations, and a viewer is asked to report the orientation. The size of the E is then adjusted based on the viewer's responses, growing smaller with correct answers or larger with incorrect answers, until the smallest E they can read is determined.

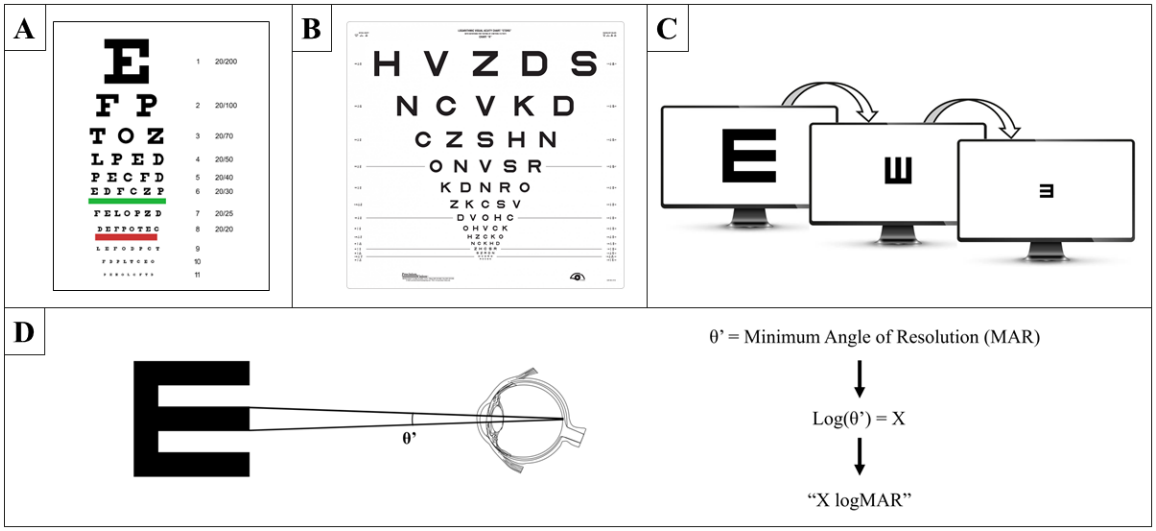


Figure 1-1: Measuring visual acuity. **A)** The Snellen chart. **B)** The ETDRS chart. **C)** A digital "Tubmling E" procedure. A viewer reports the orientation of the E on the screen, after which the E disappears, reappearing after a short delay. With each response, the size is adjusted and orientation is chosen at random from four possibilities (up, down, left, or right). **D)** How logMAR is calculated. First, the smallest feature of a letter that can be resolved is measured. The smallest feature is often taken to be $1/5^{th}$ the size of the smallest letter that can be read, as shown on the left. The angular size of that smallest feature, in arcminutes (θ'), is the Minimum Angle of Resolution (MAR). The result of taking the logarithm of the MAR is the logMAR value, as shown on the right.

VA is frequently described using Snellen notation, with 20/20 typically serving as a reference point for normal vision. A Snellen acuity of 20/40 indicates that the person would need to stand 20 ft away from an object to resolve the details that a person with normal vision could see from 40 ft away. The International Classification of Diseases and Related Health Problems sets the threshold for LV at 20/40 with the best possible correction, and the threshold for legal blindness at 20/200 (Crossland et al., 2021; Chan et al., 2018). In research contexts, VA is often given in logMAR notation. MAR refers to the "Minimum Angle of Resolution," or the angular size (in minutes of arc) of the smallest visual detail that can be resolved. LogMAR is then calculated by taking the logarithm of the angular size, as shown in Figure 1-1, Panel D. 0 logMAR converts to 20/20 Snellen acuity, 0.3 logMAR converts to 20/40, and 1 logMAR converts to 20/200.

Contrast sensitivity (CS), or the ability to distinguish between shades of grey, is another important metric for defining what can be seen in LV. CS is often measured using a Pelli-Robson chart, on which letters at the top are black against white, but progress to lighter and lighter shades of grey as one reads down the chart. The Tumbling E procedure can be used to measure CS as well, but instead of adjusting size, the contrast between the E and the background is adjusted until the lowest visible contrast is reached. The Pelli-Robson chart and the Tumbling E (for CS measurement) are both illustrated in Figure 1-2, Panels A and B, respectively. In both methods, contrast between the letters and the background is quantified with Weber contrast, calculated as the luminance of the letter minus the luminance of the background, divided by the luminance of the background. In research contexts, CS is typically reported as logCS, or the logarithm of

the reciprocal of the minimum contrast difference that can be detected. The process for calculating logCS is shown in Figure 1-2, Panel C.

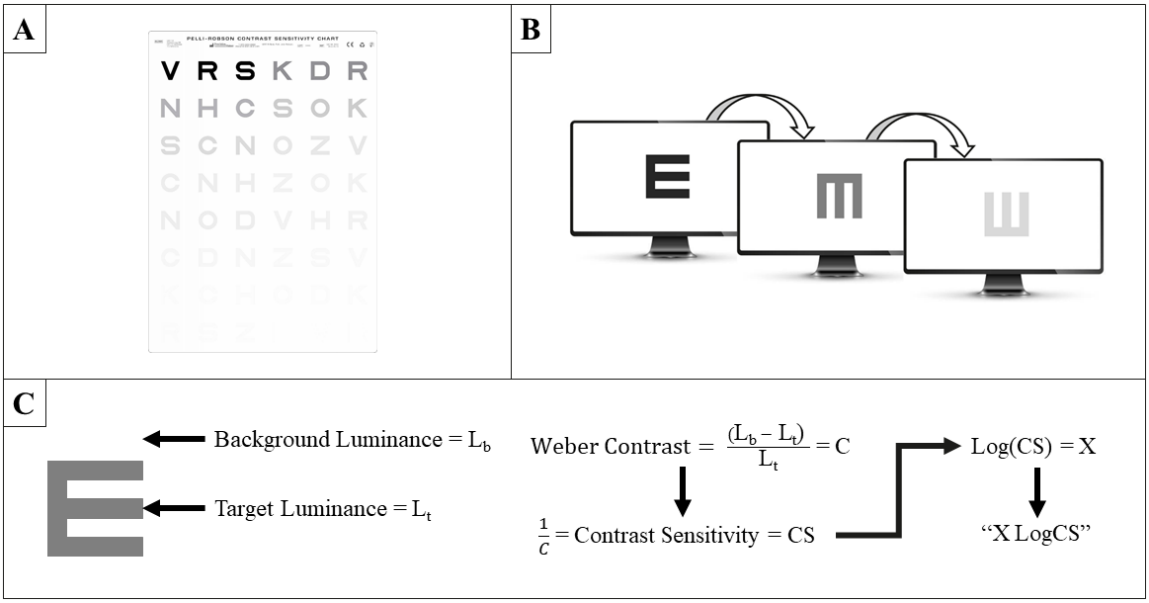


Figure 1-2: Measuring contrast sensitivity. **A)** The Pelli-Robson chart. **B)** The "Tubmling E," used to measure contrast sensitivity. The procedure is the same as for acuity measurement, except contrast of the E with the background is adjusted, instead of its size. **C)** How logCS is calculated. The lowest level of contrast between the letter and the background at which the letter can still be seen is determined, and recorded in Weber contrast. LogCS is calculated as the logarithm of the reciprocal of that level of Weber contrast.

This dissertation focused on visual search for environmental objects with low vision, and with the online measurement of VA and CS in low vision. It is comprised of three studies, described in Chapters 2, 3, and 4, one of which focuses on functional vision, while the other two focus on the limits of what can be seen. In Chapter 2, I studied visual functioning by evaluating how lighting could be used to make obstacles or other features of the environment easier to see. In chapters 3 and 4, I studied the precise quantification of what visual details can be seen with LV, describing the process of developing and validating novel measures of VA and CS that can be administered remotely.

Chapter 2: The Impact of Enhanced Lighting on Object Visibility with Restricted Acuity

Lighting is a key part of building spaces that are safe and efficient to move through with LV. Whether an object is bright enough or contrasts enough with its surroundings can make the difference between it being a helpful navigational cue or being an obstacle that presents a tripping hazard (Boon et al., 2015; Chung et al., 2018; Kuyk et al., 1996). However, more research has focused on illumination than on contrast, and while guidelines for safe levels of illumination in built spaces are available, there are no such guidelines for contrast. In this study, we conducted two experiments that focused on implementing directional lighting to make objects more clearly visible with restricted vision. We hypothesized that while additional illumination would make most targets easier to see, the lighting's impact on an object's contrast with its surroundings would be an important factor in determining how effective it was.

In both experiments, normally sighted participants were asked to search a room for target objects while wearing goggles that artificially restricted their VA to 1.8 logMAR (Snellen equivalent of 20/1250). Under typical ambient classroom lighting, participants walked along a path while visually searching for target objects adjacent to the path, and verbally reporting when they saw each object. Performance was quantified for each target with "detection rate" – the percentage of participants that reported seeing an object. The targets were white or grey Styrofoam boxes and cylinders, and the room had white walls and black doors. Figure 1-3 includes a blueprint of the placement of objects for an experimental trial in Panel A, while Panels B and C show objects with and without the

blur effect of the goggles, and Panel D shows a wider view of the room from the starting point on the path, including examples of the target objects. Spotlights were used to enhance the illumination of some of the targets to three fold their ambient illumination, which we took as an upper bound for the plausible additional illumination provided by a typical lighting intervention.

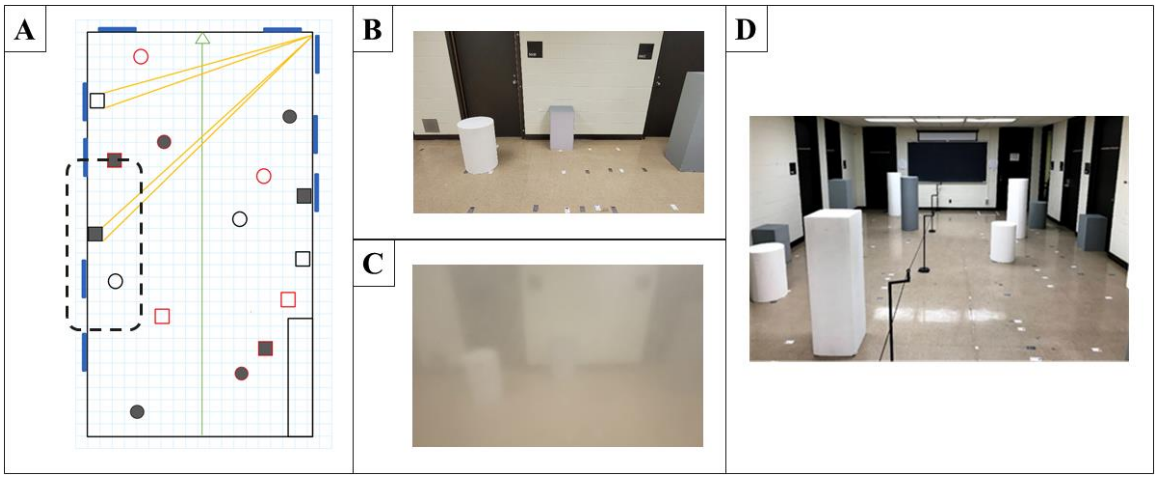


Figure 1-3: The testing environment used in the spotlighting studies. **A)** An example blueprint of one of the trials, indicating where objects were placed (squares, circles), and where spotlights were shone (yellow lines). The dotted black line indicates the area of the room photographed in Panels B and C. **B)** A grey target illuminated by a spotlight, placed against a white wall. **C)** The same target from Panel B, but viewed through the blur goggles participants wore during experimental trials. **D)** A broader view of the experimental space (with the same layout illustrated in Panel A) from the perspective of the starting point for a trial.

Spotlighting was leveraged differently in each experiment. In Experiment 1, we explored whether increasing a target's luminance, a measure of how much light is reflected off its surface, would reliably make it easier to see, even if it was difficult to see under ambient lighting. In Experiment 2, we aimed to evaluate the role of contrast in addition to luminance, so we used the spotlights to systematically vary targets' contrast with the background surfaces behind them. By combining different target colors and background colors, spotlight illumination increased contrast for some targets, while

having little impact on or decreasing contrast for others (see Chapter 2, Figure 2-5 for more detail).

We found that in both Experiments 1 and 2 that while increasing luminance helped make some targets brighter, the improvement was not consistent. Specifically, the increased luminance improved detection when it also served to increase contrast. Those cases included white targets against white walls, for which detection rate improved by 21% with spotlighting, and grey targets against black doors, for which detection rate improved by 32%. On the other hand, when increased luminance also reduced or had little impact on contrast, it had no effect on detection rates.

Results from Experiment 2 were synthesized into a set of 4 principles of luminance and contrast, which describe how both factors affect target visibility, as well as a set of 3 guidelines for spotlighting, which describe how directional lighting can be effectively leveraged to improve visibility. These principles and guidelines could serve as valuable references for parties aiming to improve the visual accessibility of indoor spaces with the use of directional lighting.

Chapter 3: Development and Initial Validation of the Online Vision Test in Normal Vision

Monitoring changes in VA and CS is valuable for tracking the progression of eye disease, low vision rehabilitation, and vision research. However, costly travel to clinics for measurement can deter some patients. While remote tests are available, many patients have sight too poor to be measured with them. This study aimed to address those issues

by developing and conducting a preliminary validation of online VA and CS tests that can both be administered remotely and to those with sight too poor to be measurable with currently available remote tests, including both online tests and smartphone apps. These tests were called the online visual acuity (o-VA) and online contrast sensitivity (o-CS), and together they constitute the Online Vision Test (OVT).

The poorest VA that can be measured with currently available remote VA tests is 20/400. However, even sight with VA poorer than 20/1600 is often still useful in daily functioning (Jeter et al., 2017). The o-VA was developed to fill that gap, and can measure acuities as low as 20/5000 on a small iPad screen, with larger displays capable of measuring even poorer acuities. The test is also simple to take, utilizing a Tumbling E procedure, as shown in Figure 1-1, Panel C. Furthermore, the o-VA is easy to calibrate, requiring only common household items of standard size for viewing distance calibration, such as a paper towel roll, or a sheet of printer paper.

A major challenge for remote CS testing on digital displays is ensuring the contrast of the stimuli is consistent across different screens. This difficulty arises because different screens (even those produced by the same manufacturer) often have differences in their gamma curve calibration, the function that relates grey levels to pixel luminance. Thus, an image encoded to display a grey stimulus against a white background will result in different luminance contrast levels on different displays. The o-CS made use of dither patterns to create a range of contrast levels that are robust to differences in gamma curve calibration. Similar to the method used by newspapers to print greyscale images, in which ink is sprayed onto paper at different concentrations of dots per square inch, dither

patterns are images composed of different proportions of white and black pixels to create a range of grey levels. An example of how dithering can create greyscale images is shown in Figure 1-4, Panel A. When presented to a viewer with low enough VA, the individual pixels are not resolved, and are instead perceived as a uniform grey surface. These grey surfaces were then used to create the stimuli for the o-CS (which utilizes a Tumbling E procedure), as shown in Figure 1-4, Panel B. As the dither patterns do not include any grey pixels, their luminance output is determined by the proportion of white and black pixels, rather than the gamma curve calibration. This allows the o-CS to display consistent contrast levels on any screen, regardless of gamma calibration.

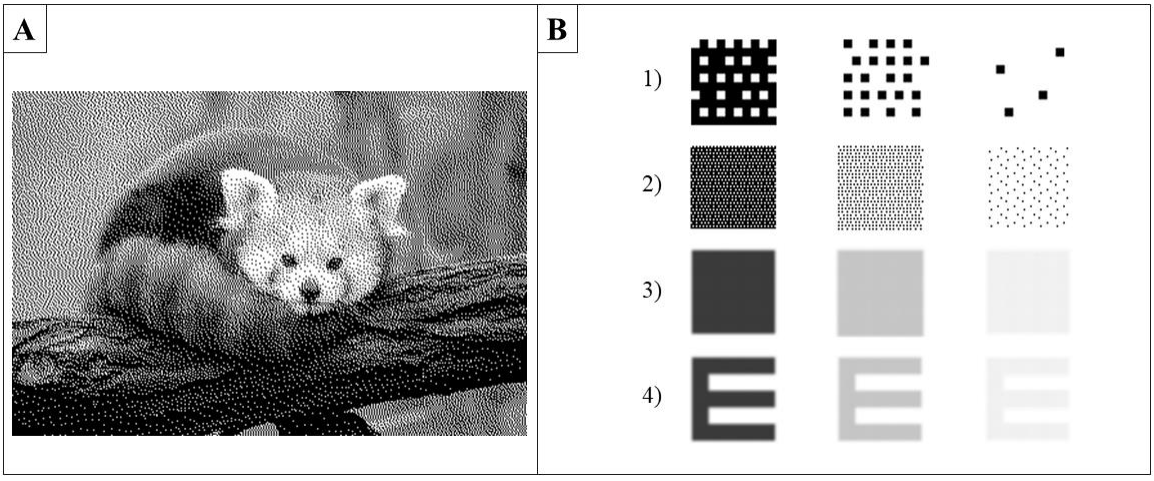


Figure 1-4: A) An example of dithering – a greyscale image is composed of only white and black dots. **B)** Dither patterns used in the o-CS. Row 1 illustrates a zoomed in image of the dither patterns, which were composed of 10 x 10 grids of white and black pixels. Row 2 shows the patterns more zoomed out, arranged in a grid, and row 3 shows those patterns as they would be seen by an observer with VA too poor to resolve the individual pixels. Row 4 shows those patterns being used to create Tumbling E stimuli.

The o-VA and o-CS were administered to participants by first running the testing software on an administrator's computer, then sending them to participants' computers via screen share. We validated the tests by comparing the results from the online tests to

results from standard clinical acuity (ETDRS) and CS (Pelli-Robson) charts in 29 participants with artificial visual restriction.

We found that the mean scores were consistently similar, but slightly better, on the online tests compared to the charts. Scores on online tests were on average approximately 0.1 log unit better than their chart counterparts, equating to 1 line on the ETDRS chart or two letters on the Pelli-Robson chart. That finding indicated that the OVT accomplishes its primary goals. It produces results that are in good agreement with gold standard clinical measures, the ETDRS and Pelli-Robson charts. The tests can be administered online, requiring only access to a computer by the test takers. Furthermore, the o-VA can measure VA poorer than 20/400, the lower bound for all other currently available remote tests. Lastly, the test is simple and easy to administer, and results are straightforward to interpret.

Chapter 4: Remote Validation of the Online Vision Test in Low Vision

We followed up our initial validation from Chapter 3 by evaluating the efficacy of the test in a more realistic use case: with remote at home testing of participants who had naturally occurring low vision. Our prior testing was done in a laboratory environment, meaning it did not include the typical lighting variation that occurs in home environments, which can affect both VA and CS measurement. Our prior participant group all had identical vision restriction (using the same blur goggles), so we aimed to determine if a group with a heterogeneous array of diagnoses would continue to show good agreement between the OVT and clinical chart-based measures. This study also

investigated the practical aspects of remote testing, including whether home testing would make calibration or test set up challenging, or whether an older participant group would be uncomfortable with the technology used in the tests.

We tested 14 low vision participants remotely with the OVT, while they were at home. o-VA and o-CS scores were then compared with scores from standard clinical chart-based measures, the Near ETDRS for VA and the Mars chart (a portable chart similar to the Pelli-Robson chart) for CS. Chart testing was conducted by an experimenter, who traveled to participants' homes to administer the tests.

Results showed that mean scores were similar between the OVT and the charts on both the VA and CS tests. This indicated the remote testing environment did not introduce any bias between the OVT and the charts. On the other hand, compared to our results from Chapter 3, there was a higher degree of variance in the differences between the OVT and the charts. This indicated that an individual's OVT scores were slightly less predictive of their chart scores. Notably, however, this high variance was driven primarily by a small number of participants with very high differences between the OVT and the charts. Removing those outliers from the dataset resulted in similar variance as was observed in the sample from Chapter 3.

With regards to the practical aspects of test administration, we found that remote testing did not introduce any additional challenges compared to testing a participant in the same room. While three participants needed the assistance of a sighted friend or family member to complete the calibration process, it otherwise proceeded smoothly

without any unforeseen challenges. Furthermore, tests were typically completed quickly (<5 minutes), and participants reported no confusion with test taking process or discomfort with the technology used in the tests.

Our findings suggest that compared to testing in a laboratory environment, remote testing at home led to slightly poorer agreement between the OVT and the charts. While it did not introduce any bias to mean OVT scores results relative to chart scores, but it may have introduced greater variance in the differences between the tests. As this increased variance appears to be attributable to a small number of outliers, additional data will be necessary to draw definitive conclusions about the agreement between the OVT and the charts. That being said, the OVT shows promise, demonstrating the remote testing process is quick and easy for both the administrator and the test taker, and showing good agreement with gold-standard clinical charts for the majority of our participants.

Conclusion

Vision impairment affects not only what can be seen, but also the many functions that rely on visual information to be carried out. This thesis addresses both of these aspects, assessing visual function in the context of a visual search task, and assessing what can be seen in the context of developing novel VA and CS measurement tools.

The study focused on visual functioning found that directional lighting can be used to enhance the visibility of objects for observers with low vision. However, it also demonstrated that simply making things brighter is not enough to make people see them –

illumination is most effective when it is used to create or enhance contrast between objects and their surroundings.

In a second line of research, I found that it is possible to effectively measure both VA and CS remotely. The OVT is capable of measuring VA to more severe visual impairment than any other currently available remote test, and can measure CS on whatever computer display is available to the test taker in their home. It demonstrates good agreement with gold-standard clinical measures for most participants and is quick and easy to use, providing a valuable remote option to test those who would otherwise be unable to visit a clinic for regular vision checkups.

Chapter 2: The Impact of Enhanced Illumination on Target Detectability with Artificially Reduced Acuity

Introduction

Independent mobility is an important part of daily life, and for people with low vision (LV), it can be enabled or hindered by the architecture and interior design of built environments. Qualitatively, low vision is any form of chronic vision impairment that impacts daily functioning and cannot be corrected with glasses. Quantitatively, the International Classification of Diseases and Related Health Problems defines LV as having a best-corrected visual acuity poorer than 20/40 (Crossland et al., 2021). Prevalence estimates suggest there were nearly 6 million Americans living with low vision in 2017, and as the country's population ages, that figure is expected to rise to nearly 10 million in the next 30 years (Chan et al., 2018). Similar trends are predicted worldwide; the number of people in the world living with moderate and severe low vision in 2020 was approximately 237 million, but is expected to rise to 588 million by 2050 – an increase of nearly 60% (Bourne et al., 2017).

While their vision is impaired, people with LV still have usable sight, and typically utilize visual cues to navigate and avoid obstacles while walking. However, some cues or hazards that may be visible for someone with normal vision may not be visible for a person with low vision. This can lead to difficulties in mobility, often characterized by an increased risk for falls or collisions (Dev et al., 2021; Elliott et al., 1995; Figueiro et al., 2011; Fuhr et al., 2007; Klein et al., 1998). Hazard risk is salient to

people with vision loss – self-reported poor vision quality is associated with a greater fear of falling, restriction of travel and social activities (Donoghue et al., 2014), and reduced confidence in one's mobility (Bibby et al., 2007). The resulting restriction of access can lead to difficulty achieving activities of daily living and, ultimately, poorer quality of life (Enoch et al., 2019; Guerry et al., 2020). One way to mitigate these issues is to design indoor spaces with visual accessibility in mind. Visual accessibility is analogous to any other form of accessibility, in that it involves building or adapting an environment to support its use by people with a disability. For a space to be specifically visually accessible, it must meet three primary criteria: 1) it must support safe and efficient travel through the environment, 2) it must assist in the perception of the spatial layout of key features in the environment, and 3) it must provide information that helps one keep track of where they are and where they are going within the environment (Thompson et al., 2022). Each of these criteria covers a different aspect of interacting with a space, but they are all united by a key goal: improving the visibility of important features of the environment, whether they are potential mobility hazards or helpful navigational cues.

Lighting and Contrast

Many factors affect the visibility of key features in a space, but illumination has received the most research attention. Particularly in the context of mobility, research has consistently demonstrated that for people with LV, low illumination impairs hazard detection and avoidance (Chung et al., 2018; Fuhr et al., 2007; Geruschat et al., 1998; Grech et al., 2017; Kuyk et al., 1996), as well as leading to slower walking speed (Kumaran et al., 2020; Kuyk et al., 1998a; Kuyk et al., 1998b; Kuyk et al., 1999).

Another important factor in the visibility of key features is luminance contrast, or the ratio of luminances between a target and its background. Research has shown that contrast is also important for hazard detection, with lower contrast hazards being more difficult to avoid with LV (Hassan et al., 2007; Kuyk et al., 1996). Other work has focused on specific features in the environment, such as a study by Manandhar et al. (2022), which examined the visibility of door frames and handles, staircases, and light switches. By presenting images digitally, they were able to systematically vary the contrast of these building elements with their surroundings, and asked normally sighted and LV participants to rate how visible they were. They found that while normally sighted participants required only 15 – 20% contrast to rate these features as highly visible, participants with moderate LV required 15 – 48%, and participants with severe LV required 33 – 69% contrast. Similarly, other studies have asked LV participants to visually locate and rate how visible tactile ground surface indicators (which are frequently used to warn pedestrians that they are approaching a street crossing) are. These studies found that indicators with high contrast with the surrounding surface are more likely to be seen and are also rated as more visually conspicuous (Jenness & Singer, 2008), particularly for those with severe LV (Lukman et al., 2020). Beyond laboratory studies, people with LV report that luminance and contrast in the environment are important. A study by Boon et al. (2015) asked LV participants to keep a diary of situations in which they experienced injurious incidents or near misses that were related to their vision problems, and what they attributed those incidents to. While several factors emerged, among the top were objects not standing out due to low contrast with the

background and insufficient lighting. Taken together, these studies support the notion that both illumination and contrast are important design considerations when creating visually accessible spaces.

Lighting Guidelines for Low Vision

In the architecture and design industries, guidelines exist to inform designers and architects about the importance of lighting, as well as to aid in the development of illumination and contrast regulations. A report published by the National Institute of Building Sciences (NIBS) in 2015, titled *Design Guidelines for the Visual Environment*, emphasizes the importance of quantified lighting guidelines for visual accessibility, and provides a framework for understanding how accessibility problems might arise at different parts of a building (e.g. lobbies, stairwells, corridors, etc.). The NIBS guidelines follow this up with some recommendations for specific illumination and contrast guidelines in these locales, but also stress that many of them have not been tested, and will require data-driven evaluation before being considered definitive. The Illuminating Engineering Society (IES) subsequently released *ANSI/IES RP-28 – Lighting and the Visual Environment for Senior Living* in 2016, which provides additional, more specific guidelines for appropriate illumination and contrast levels specifically oriented toward low vision.

Unfortunately, despite the available research suggesting the importance of illumination and contrast in the available guidelines, the application of those guidelines in practice is not consistent. While the recommendations from the IES are recognized both

in the United States and in other parts of the world as valuable (Hegde et al., 2017), they are not officially part of the Americans with Disabilities Act (ADA), and are thus not legally binding. They are referenced in some legal documents, such as the Department of Health and Human Services' *Requirements for Long-Term Care Facilities*, which outlines regulations for facilities designed specifically for older people to live in, but only as a reference for those who wish to seek more information. Instead, those requirements frequently refer to "adequate and comfortable lighting levels in all areas," and "use of contrasting colors," without specific luminance or contrast values. This language is similar to what is found in the ADA, which relies on regulations such as "detectable warning surfaces shall contrast visually with adjacent walking surfaces," and "higher lighting levels at elevators help to locate the elevator." As a result, studies of lighting in nursing homes have demonstrated that many are frequently dimly lit and without proper contrast cues, leading to poor visual accessibility in spaces specifically designed for people who are likely to experience LV (Hegde et al., 2017; Sinoo et al., 2014). Furthermore, there is little available published research on the actual efficacy of the IES guidelines, so even if they were implemented properly, while it would likely be helpful for accessibility, the degree to which it would help is unclear.

Illuminance and Luminance

Another issue with existing guidelines is that they are typically illuminance based, rather than luminance based. Illuminance is a measure of the total luminance flux (i.e. how much light) is in the room, but not how that light interacts with the surfaces within the room, whereas luminance is a measure of the amount of light reflected from a specific

surface in a specific direction (Thompson et al., 2022). The predominance of illuminance-based guidelines presents three issues. The first relates to viewers within a space: illuminance does not on its own describe how objects in the space will be seen, as not all the light present in a space will actually reach the eyes of people in that space. Luminance, on the other hand, incorporates a viewpoint in the measurement, meaning luminance-based guidelines can more accurately reflect what viewers see. The second issue is related to objects within a space: illuminance based guidelines do not address how lighting affects an object's contrast with its surroundings. If an entire space is illuminated, it may do little to improve an object's contrast with its surroundings, as it increases the amount of light on both the object and the background behind it. On the other hand, as luminance describes light in relation to specific surfaces, guidelines that utilize luminance values can more effectively describe specific levels of object-background contrast. The third issue is related to the cost associated with illuminance-based guidelines: meeting illumination thresholds for entire indoor spaces can be quite resource intensive. Modern designers face increasing pressure to improve energy conservation, and the increased cost associated with illumination guidelines can be a barrier to their effective implementation (Kersten et al., 2013). Luminance-based guidelines, which would focus important lighting thresholds on areas where they are most impactful, could be more cost effective and thus plausible to reliably meet.

Directional Lighting and the Current Study

There is a potential alternative to broadly illuminating entire spaces, one which is more directed and strategic. Directional lighting, the focused application of illumination

towards a specific location, could be used to effectively meet luminance-based lighting guidelines. Directional lighting has also been shown to improve 3D shape recognition under conditions of visual impairment (Kallie et al., 2012), which may help in recognizing and utilizing navigational cues. Additionally, focused illumination can be used to effectively create luminance contrast boundaries, which can be useful for highlighting important features of the environment. Aiming to leverage this approach, the present study was conducted to determine if directional lighting could be used to improve object visibility under conditions of restricted vision. To investigate that hypothesis, we conducted a pilot study (briefly summarized below) and two experiments (which are the focus of this report).

The pilot study was conducted to evaluate the visibility of different types of objects under conditions of artificial vision restriction. In the pilot study, normally sighted participants walked through a room while wearing blur goggles that restricted visual acuity, contrast sensitivity, and visual field, and were instructed to find and describe target objects placed adjacent to their walking path. Results showed three potential factors that may have influenced target visibility: the target's color (gray or white), its size (small or large), and its position along the walking path. Grey targets were seen less frequently than white targets, small targets were seen less frequently than large targets, and targets placed such that participants turned away from them while walking were detected less frequently than targets at other locations along the walking path.

In Experiment 1, we investigated whether directional spotlight illumination would enhance the visibility of targets that are difficult to see with reduced acuity. Specifically,

we aimed to determine if the reduced visibility of targets due to color, size, and positioning that were observed in the pilot study could be reliably replicated. Furthermore, we also aimed to determine whether spotlight illumination would increase the likelihood that participants would see those targets. We hypothesized that the factors which reduced visibility in the pilot study would do so again under ambient overhead lighting, but that those same targets would be seen more frequently when they were illuminated by spotlights. This hypothesis was investigated by asking participants to walk through an indoor space while wearing goggles that restricted vision, but with the addition of systematic lighting variation on the targets that were hypothesized to be more difficult.

Experiment 2 was conducted to expand on findings from Experiment 1, namely that some of our particularly difficult targets benefited from directional spotlight illumination while others did not, and is the focus of the majority of our analysis. When considering our findings from Experiment 1, we carefully examined our experimental space and found that under some circumstances, increasing a target's luminance increased its contrast with the background surface behind it, but under others, the increased luminance actually reduced the target-background contrast. This led us to ask the primary question of Experiment 2: is a target's contrast with the surface behind it as important as the target's luminance in determining its visibility? Experiment 2 utilized similar methodology to Experiment 1, but contrast between spotlight-illuminated targets and the background behind them was more systematically controlled, such that spotlight illumination could increase, decrease, or have little effect on contrast. We hypothesized

that the combined effects of increased luminance and changes to contrast would determine whether enhanced lighting will benefit detectability, meaning the interplay between these factors will be of major importance to implementing lighting for visual accessibility.

Methods

Materials, Stimuli, and Search Task

Both Experiment 1 and 2 utilized many of the same materials and stimuli, and the search task was also identical in the two experiments. Those shared materials, stimuli, and task are explained here first, followed by features that were specific to each experiment in the Experiment 1 and Experiment 2 segments of the Methods section.

Artificial Acuity Restriction Goggles

Figure 2-1: The goggles used to artificially restrict participant's acuity. Peripheral viewing was completely blocked except through the Bangerter Foils.

Participants wore goggles which reduced their acuity to approximately 1.8 logMAR (Snellen equivalent of 20/1250) and their contrast sensitivity to approximately 0.5 logCS, shown in Figure 2-1. Additionally, the viewing aperture of the goggles limited

their visual field of view to a circular window of approximately 70° in diameter. The goggles were opaque, aside from a single aperture over one eye (limiting participants to monocular viewing). A circular plano lens (a piece of plastic with no refractive power) was secured in the aperture, and Bangerter Occlusion Foils (Odell et al., 2008) were adhered to the lens to produce blur (this method was also used in previous work by our lab to restrict acuity, Legge et al. 2010; Freedman et al., 2018).

Targets and Backgrounds

The primary task during each trial was to visually search for and identify target objects that were placed adjacent to the walking path. Target objects were made from polystyrene, and varied in color, size, and shape. Colors included white and grey, sizes included short and tall (2 and 4 ft in height, but always 1 ft in diameter), and shapes included boxes and cylinders. Figure 2-2 shows two photographs of what the targets looked like. Two of each type of object were used for each trial, for a total of 16 targets per trial. Targets were placed at different locations for each trial, requiring participants to

Figure 2-2: Photos of the targets used in the experiment. On the left, targets are set up for an experimental trial, such that both tall and large, white and grey, and boxes and cylinders are visible. On the right, a front-on shot which offers a clear height comparison between tall and short targets.

search a novel environment in each trial. Targets were placed such that they did not obstruct the path participants walked on.

The walls of the room, forming the background for the objects, were made primarily of two kinds of surfaces: walls of brick painted white, and doors painted black. The range of luminances of both of these surfaces are included in Table 2-1.

Illumination Conditions

Two illumination conditions were used throughout both Experiments 1 and 2: ambient overhead lighting, and directed spotlight illumination. The ambient overhead lighting served as the default lighting condition. The directed spotlight served as an enhanced lighting condition, added in addition to the ambient lighting, and was directed only at targets designated for lighting enhancement. The possible range of luminance values for targets under each illumination condition are displayed in Table 2-1.

Ambient lighting was produced by four rows of fluorescent bulbs on the ceiling, which remained on for the entire experiment. All targets were illuminated by the ambient lighting throughout all trials, and a subset of those targets received additional illumination from the spotlights. Overhead lights were spaced uniformly across the ceiling, with eight feet between each row. The mean illuminance in the room under ambient illumination was 400 lux, fluctuating between 275 and 600 lux, depending on proximity to overhead lighting fixtures.

Directed spotlight illumination was produced by ETC brand Source Four spotlights with 750-watt incandescent bulbs, and was used to increase the luminance of

two targets per trial. The spotlights were positioned in small side rooms adjacent to the main room in which the experiment was conducted. The spotlights were shone into the main room, such that participants could not see the spotlights during trials, and the light was never directed above waist height of any given participant. As a safety precaution to prevent participants from looking directly into the beam of the spotlights, participants were informed that spotlights would be used to illuminate certain targets. Participants were told that they may notice a bright light at the edge of the room, but were told not to look directly at it if they noticed it.

Under direct spotlight illumination, enhanced targets received extra lighting proportional to their ambient luminance. With spotlight illumination, the luminance of an enhanced target was 3 (±.1) times greater than with ambient lighting. Therefore, the absolute difference between a target under ambient illumination and one under direct spotlight illumination was dependent on the luminance of the target under the ambient condition.

Surface	Luminance (cd/m^2)
White Target - Ambient Illumination	70-140
Grey Target - Ambient Illumination	30-38
White Target – Enhanced Illumination	210-420
Grey Target – Enhanced Illumination	90-114
White Wall	80-120
Black Door	3-5

Table 2-1: Luminance values for targets under both lighting conditions and for background surfaces.

Only one side of the target, the one facing the path, was illuminated. To do this, the beam of light needed to originate from the opposite side of the room, and it was therefore necessary for the beam to cross the walking path. This meant that participants

would at some point walk through the beam and occlude the light on the target, causing it to darken in the participant's shadow before brightening after they continued past the beam. The impact of this temporary occlusion was minimized by positioning the spotlights such that participants would pass through the beam after they had already passed the illuminated target. No participants reported noticing the change.

Box shaped targets were the only targets to receive enhanced illumination to avoid introducing a luminance gradient associated with shining a spotlight on a rounded surface of a cylindrical object, which prior research has shown can be a helpful cue in shape recognition tasks (Kallie et al., 2012).

Walking Path

In both experiments, participants began each trial on one end of the room, then walked to the other end along a predetermined path, and when they reached the end of the path, the trial was concluded. Participants were instructed to not reverse walking direction, and changes in walking direction did not occur. The path was delineated by a thin rope suspended on stanchions, which served to guide them safely through the room, and provided a cue if they needed to make a turn on the path. Participants were instructed to hold onto and follow the guideline at all times during experimental trials.

While the same basic materials were used for the path across both experiments, the two experiments used a slightly different path through the space. In Experiment 1, two symmetrical paths were used, illustrated in Panels A and B in Figure 2-3 as blueprints. The paths were 15 m (49 ft) in length, and included four 90° turns, 2 to the left

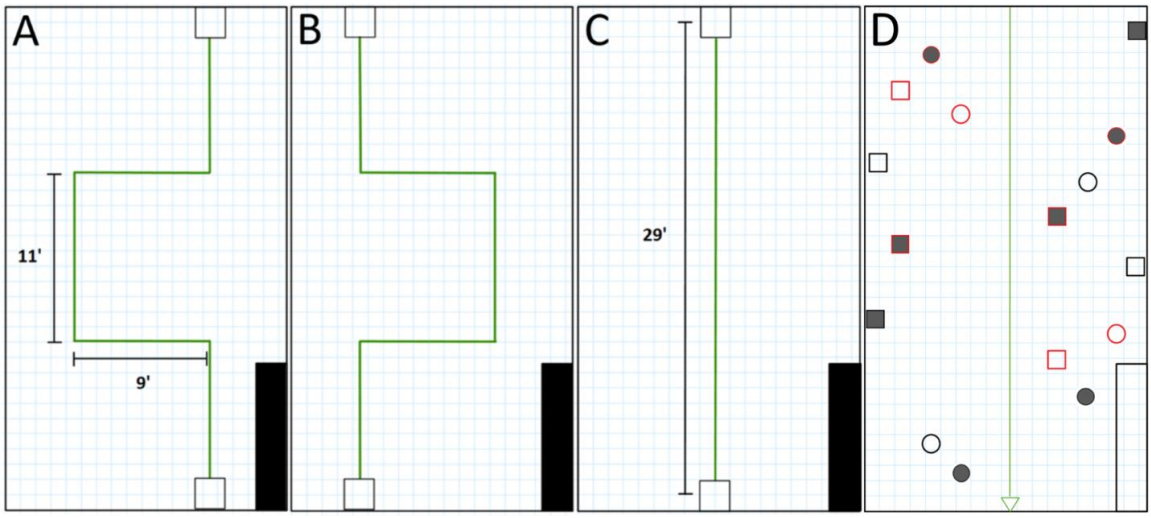


Figure 2-3: Blueprints of the experimental space and the paths participants walked along. The green line indicates the path that participants walked along, and the boxes at each end indicate starting/stopping points. Each square on the background of the blueprints is scaled to one square foot. Panels A and B illustrate Path 1 and 2, respectively, from Experiment 1, while panel C illustrates the path from Experiment 2. Panel D is an example of the blueprint with target locations and participant walking direction included, illustrating target positions relative to the path with boxes and circles, and walking direction with the green arrow.

and 2 to the right. In Experiment 2, one path was used, illustrated in Panel C of Figure 2-3. The path was approximately 9 m (29 ft) in length, and included no turns. To determine if there was an effect of walking direction, half of the trials began on one end of the room, and half on the other end (starting points are indicated by the boxes in Figure 2-3) in both experiments. Targets were placed in similar proximity to the path in both experiments, illustrated in Figure 2-3, Panel D.

Search Task

Before beginning any experimental trials, participants were given instructions about how to complete the search task. After they affirmed their understanding of the instructions, a single practice trial was conducted before the main experiment began. This

allowed participants a chance to become accustomed to the restricted vision through the goggles, and introduced them to what targets would look like with restricted acuity. Only 3 targets were placed in the room during practice trials, but participants were asked to report their size and shape as they would on a real trial, and to walk the full length of the path. After practice was complete, participants exited the room while the experimental trials were set up.

Before entering the room for experimental trials, participants were instructed to wear the blur goggles, at which point they were guided by an experimenter to the trial's starting location while keeping their gaze directed downward toward the floor, to avoid preemptive viewing of the room. After a countdown to begin each trial, they walked through the room at whatever pace felt most natural to them. Participants were instructed to search the room for targets while walking, and to verbally report their size and shape once they saw one. Specifically, they were instructed to first report the size as either "tall" or "short," then to report the shape as either a "box" or a "cylinder." Alternatively, if they were unsure about either the size or the shape, they were allowed to answer that they were unsure about either or both properties. If a participant gave any response to a target, whether their size and shape were correct or incorrect, or if they reported being unsure, their response was coded as a detection. If they did not report anything about a target, it was coded as a miss. The percentage of participants who detected a target was that target's detection rate, which was used as the primary outcome measure for both Experiments 1 and 2. Notably, the definition of "detection" used in this study is a departure from the more traditional definition used in earlier in psychophysical research,

which would typically refer to whether or not a target could be distinguished from no target. Our use of detection differs in that targets are always in the room, meaning the task is focused more on searching for targets, rather than determining their presence or absence. Additionally, while both size and shape recognition data were recorded, both showed a strong ceiling effect, and were thus not used for subsequent analysis.

During the search task, participants were instructed to use their free hand to point at the detected target to ensure the experimenter knew which target was being reported. They were told that they could stop for a moment to point at the target, but to not spend a prolonged length of time on any one target (no more than one or two seconds). Trials ended when the participant reached the end of the path. After each trial, participants were asked to step into an adjacent room while the next trial was set up.

Experiment 1: Applying Spotlight Illumination to Difficult Targets

Participants

25 normally sighted participants participated in Experiment 1, with a mean age of 24 (range = 19-38 years), nine of which were female. All participants were normally sighted, with visual acuity that was self-reported to be corrected to 20/20 acuity or better. All participants gave written informed consent after the nature of the study was described. The research procedures were approved by the University of Minnesota Internal Review Board, and adhered to the tenets of the Declaration of Helsinki.

Trials and Blur Conditions

Participants in Experiment 1 completed 1 practice trial and 10 experimental trials. The first nine experimental trials were completed while wearing the blur goggles, while the final trial was completed without the blur goggles. The final, unrestricted trial provided a within-participant control to determine if participants would have difficulty completing the task without the visual restriction. Participants typically detected all targets on unrestricted trials (less than 2% of targets were missed across all participants on unrestricted trials), and as such, no further evaluation of unrestricted trials will be reported here.

Enhanced Targets

We hypothesized that three kinds of targets would be particularly difficult, and aimed to determine whether they would be seen more frequently with additional lighting compared to ambient lighting. Therefore, we had two principle independent variables in Experiment 1: target type, and lighting condition. The three types of targets were targets with grey color (GC), targets that subtended small visual angles (sVA), and targets that were placed such that participants would turn away (TA) from them at one of the turns in the path. GC targets were always tall grey boxes, and were placed such that when a participant passed them, it was on a straightaway on the path. sVA targets were always short white boxes, and were placed as far away as possible from the path (13 ft), against the wall of the room. Turn Away targets were always tall white boxes, 0.5 ft away from the path, but placed at a turn in the path, such that a participant's egocentric motion at the

point of nearest approach was turning 90° away from the target. This positioning is illustrated in Figure 2-4.

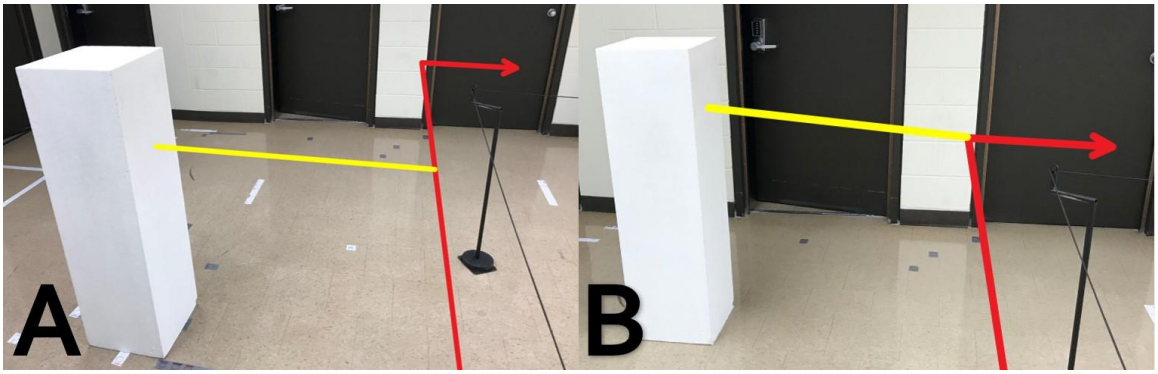


Figure 2-4, Turn Away targets: The red lines in both panels A and B illustrate the path a participant would be walking along, while the yellow lines illustrate a participant's line of sight at the point along the path closest to the object, where they are most likely to detect it. Panel A shows a non-TA target, placed such that participants would be walking in a straight line when they passed it. Panel B shows a TA target, placed such that participants are turning away from its location as they follow the path.

Each trial included two of each type of difficult target, under either spotlight or ambient illumination. In each trial, two targets were enhanced with spotlight illumination, while all other targets were under ambient illumination. Across all nine experimental trials, each of the three difficult target types were under enhanced spotlight illumination six times, and each was under ambient illumination 12 times. To facilitate comparison between those conditions, when they were presented under ambient illumination, targets were placed in the same location as when they had been spotlight-illuminated in other trials.

Experiment 2: Impact of Luminance Level and Contrast Level

Participants

31 normally sighted participants participated in Experiment 2, with a mean age of 31 (range = 19-58 years), 19 of which were female. All participants were normally sighted, with visual acuity that was self-reported to be corrected to 20/20 or better. All participants gave written informed consent after the nature of the study was described.

Trials

Participants in Experiment 2 completed one practice trial and 10 experimental trials. All trials were conducted with the participants wearing the blur goggles, though participants were familiarized with what target objects looked like with and without the goggles during explanation of trial procedures.

Enhanced Targets

Targets that received spotlight illumination were all small boxes, colored either white or grey, placed directly against the walls of the room, colored either white or black, at a distance of 9 ft from the travel path. The short boxes were the targets to receive spotlight illumination because they tended to be the least frequently detected targets in Experiment 1, meaning they had the most potential for improvement from directional lighting. Targets were placed directly against walls to minimize changes in target-background contrast that would have been due to changes in viewpoint as participants walked along the path. This setup allowed for four target-background combinations: grey-target x white-wall, grey-target x black-door, white-target x white-wall, and white-

target x black door. Additionally, there were two illumination conditions for these targets: ambient and enhanced with spotlight illumination. These factors combined to create a total of eight target x background x illumination conditions. Examples of each of these combinations is pictured in Figure 2-5, along with typical contrast values (calculated in Michelson contrast) between the target and its background.

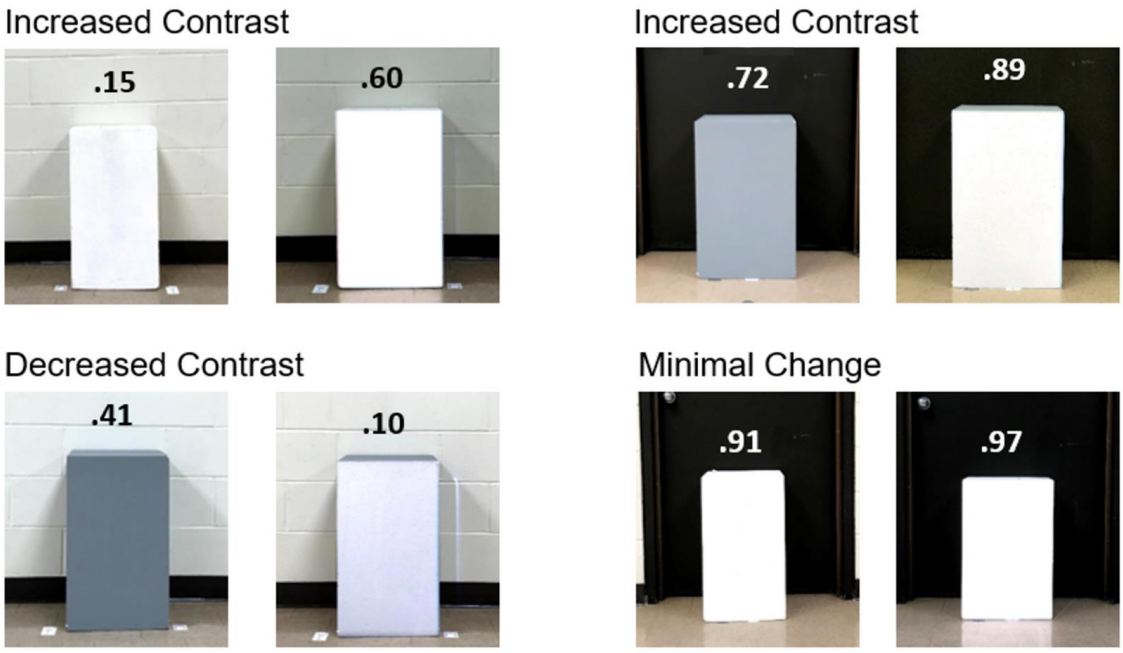


Figure 2-5: Photos illustrating each of the target color x background color x illumination condition combinations. Each pair of photos shows a color x background combination, with the ambient illumination on the left and the enhanced illumination on the right. Both pairs on the top row of images showcase how enhanced illumination could increase contrast. The bottom left showcases how enhanced illumination could reduce contrast, while the bottom right showcases how adding extra illumination to an already high-contrast pairing has little impact on contrast.

Under ambient illumination, grey targets had moderate contrast against both backgrounds, ranging from 0.39 – 0.54 against white walls, and from 0.70 – 0.73 against black doors. White targets had low contrast against white walls, ranging from 0.12 – 0.18, but high contrast against black doors from 0.88 – 0.91. As in Experiment 1,

directional spotlights were used to illuminate the face of the target pointing toward the path. The impact of this spotlight illumination on target-background contrast was dependent upon which background the target was in front of. For grey targets, enhanced illumination increased contrast for those against black doors to a range of 0.88 – 0.90, but decreased contrast for those against white walls, to a range from 0.07 – 0.13. For white targets, enhanced illumination increased contrast for those against white walls, to a range from 0.55 – 0.62, but had a very small increase in contrast for those against black doors, to a range from 0.96 – 0.97, as contrast was already very high.

Enhanced targets in Experiment 2 also shared two more similarities with the procedure from Experiment 1. First, in each trial, two targets received enhanced illumination, while others were all under ambient overhead illumination. Second, the objects were in the same location when they were under ambient illumination and when they received enhanced illumination.

Results

Experiment 1: Applying Spotlight Illumination to Difficult Targets

Effect of Difficult-to-Detect Target Types

We first evaluated our hypothesis that targets categorized as sVA, GC, or TA would be more difficult to detect than other targets. First, we calculated a baseline detection rate, defined as the mean detection rate for all targets that were not categorized as sVA, GC, or TA, pooled across all participants on all vision restricted trials. The baseline detection rate was 78.9% (*SD* = ±13.5%). This baseline detection rate was then

compared to the mean detection rates for sVA, GC, or TA targets under ambient illumination. The detection rates for each category were 46.3% (*SD* = ±25.3%) for sVA, 78.7% (*SD* = ±28.6%) for GC targets, and 75.3% (*SD* = ±20.2%) for TA targets. Comparisons to baseline via t-tests showed that sVA targets were the only target category detected significantly less frequently than baseline, $t(24) = 9.56$, $p < .005$, indicating that sVA targets were especially difficult to detect, but GC and TA targets were not.

Effect of Spotlight Illumination

Next, we evaluated the effect of spotlight illumination on the detection rate of sVA, GC, and TA targets. A two-way repeated measures ANOVA indicated that there was a statistically significant interaction between target type and illumination condition (ambient or spotlight illumination), $F(2,24) = 6.403$, $p = .008$, suggesting that spotlight

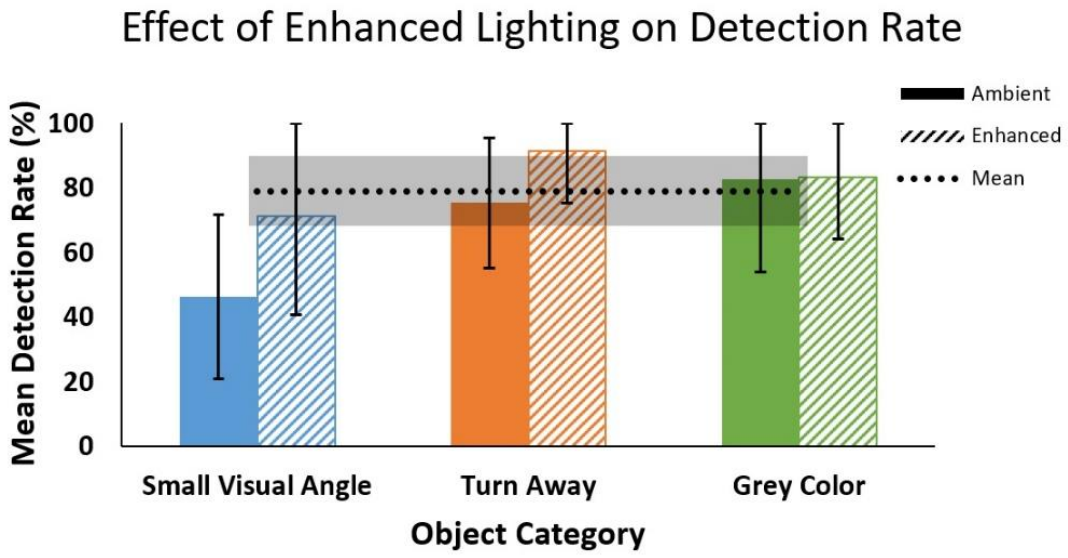


Figure 2-6: Mean detection rates across all 25 participants for each target type under ambient illumination and spotlight illumination. Error bars indicate standard deviations for detection rate for each type of target (maximized at 100%). The dotted line indicates the baseline mean detection rate, and the shaded area indicates the standard deviation of the baseline.

illumination affected detection for some target types, but not others. Detection rate increased under spotlight illumination for all categories, though post-hoc paired samples t-tests indicated the increase was only statistically significant for sVA targets (mean ambient = 46% vs mean spotlight = 71%), $t(24) = 6.09$, $p < .005$, and TA targets (mean ambient = 75% vs mean spotlight = 91%), $t(24) = 5.71$, $p < .005$. Detection rates of each type of target, under both ambient illumination and enhanced illumination, as well as the baseline detection rate for other targets, are illustrated in Figure 2-6.

Experiment 2: Impact of Luminance Level and Contrast Level

Detection Rate Descriptive Statistics

The mean detection rates for targets of each type, pooled across all trials for all participants, are listed in Table 2-2. Note that the table includes both illumination conditions and backgrounds for the short boxes.

White Targets – Mean Detection Rate

	Box	Cylinder
Short	79%	92%
Tall	91%	93%

Grey Targets – Mean Detection Rate

	Box	Cylinder
Short	50%	66%
Tall	77%	65%

Table 2-2: Mean detection rates for all targets in Experiment 2.

The remainder of the analysis will focus on the primary targets of interest in Experiment 2, the short boxes, which were the targets to receive systematic luminance and contrast manipulation via spotlight illumination. Short boxes have been analyzed in categories depending on their color, illumination condition, and the background they were placed against. Dividing the stimuli into these categories, each participant was

exposed to each stimulus configuration (i.e. the target's specific combination of color, illumination, and background) a total of 5 times, meaning that across all 30 participants, a total of 150 observations were possible for each configuration. The mean and standard deviations of these categories are listed in Table 2-3, and displayed in Figure 2-7.

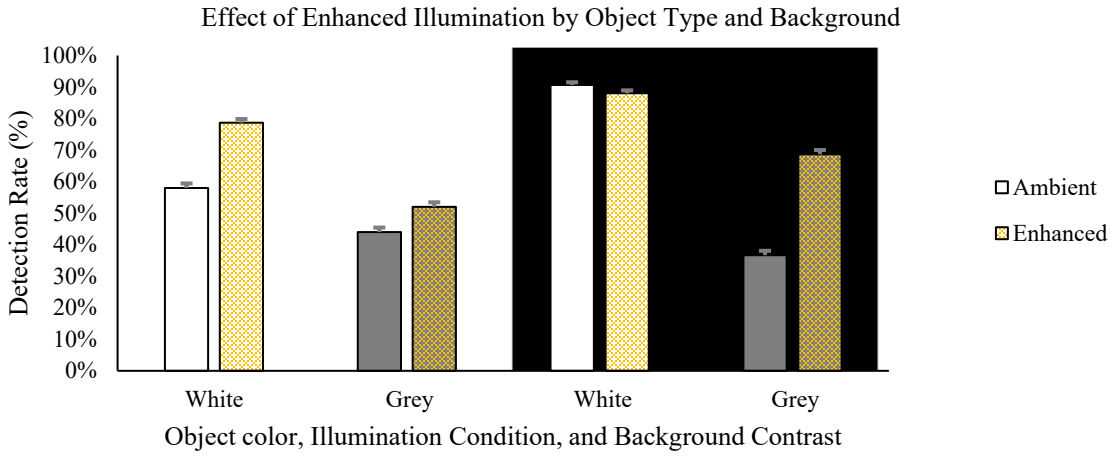


Figure 2-7: Mean detection rates for short boxes across all color, background, and illumination conditions, with error bars indicating standard deviations. White bars correspond to boxes that were colored white, while grey bars correspond to boxes were colored grey. Solid colored bars correspond to boxes illuminated by ambient lighting, while bars with yellow cross-hatching correspond to boxes enhanced by spotlight illumination. Bars on the left, aginst the white background, correspond to boxes placed in front of the white wall background, while bars on the right, against the black background, correspond to boxes placed against the black door background.

White Boxes

Background	**Illumination**	
	Ambient	Enhanced
White	58% (±2.9%)	79% (±2.3%)
Black	91% (±1.7%)	88% (±1.9%)

Grey Boxes

Background	**Illumination**	
	Ambient	Enhanced
White	44% (±2.9%)	52% (±2.9%)
Black	37% (±2.8%)	69% (±2.7%)

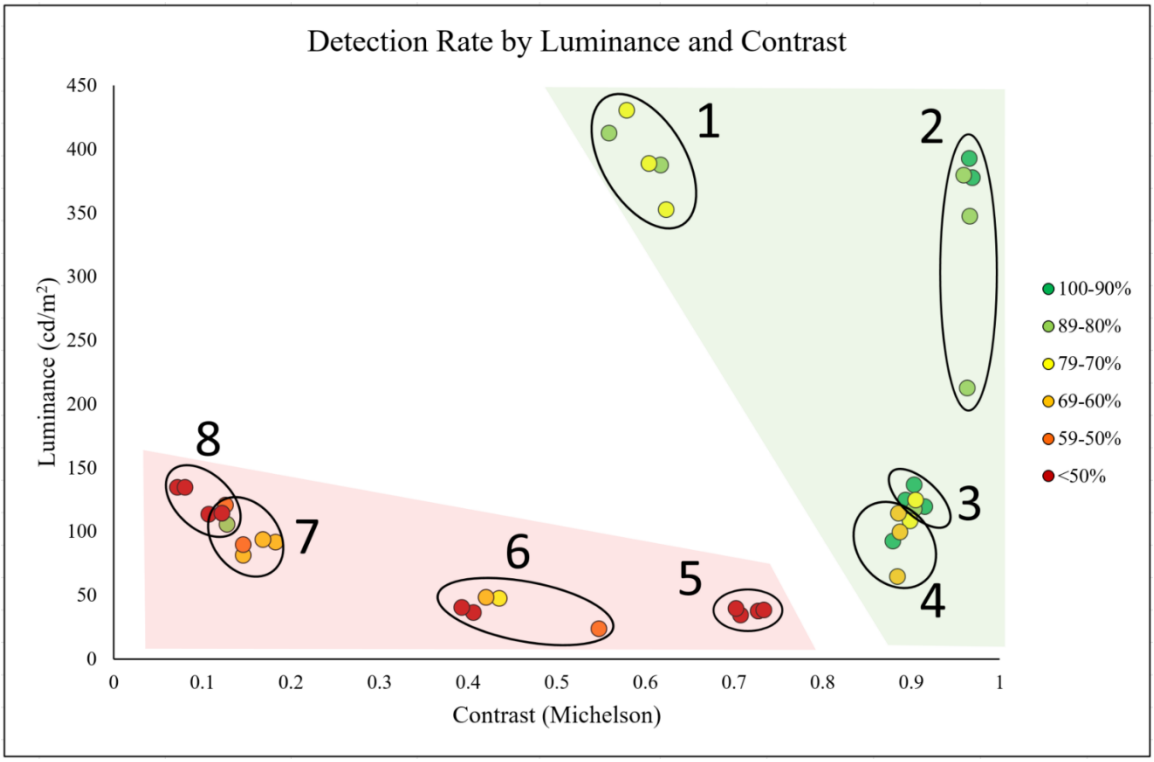
Table 2-3: Mean (±*SD*) detection rates for targets that underwent both ambient and enhanced illumination, as well as systematic manipulation of background color.

Impact of Luminance and Contrast

The results from Experiment 1 showed that enhanced illumination improved detection rate for some targets, but not all of them, suggesting that other factors in addition to luminance contribute to the detectability of targets. We hypothesized that another principle factor affecting detectability was the target's contrast with its background. To evaluate the relative impacts of luminance and contrast on detection rate, we first conducted a multilevel binary logistic regression analysis. Upon evaluating the models, we found that the models which included luminance and/or contrast had no better predictive power than a simple null model including only an intercept. However, given that prior research suggests that both contrast and luminance should have an impact on visibility (Fuhr et al., 2007; Kumaran et al., 2020) it was possible that the low predictive power may have been due to complex interactions between luminance and contrast not accounted for by the model (e.g. increasing luminance can simultaneously increase contrast, but it can also reduce contrast). To explore the possible interactions in more detail, we continued with a graphical analysis, as follows.

As shown in Figure 2-8, each target was added to a scatterplot, using their contrast as the x-axis coordinate and their luminance as its y-axis coordinate. The color of each point indicates the target's detection rate, with greener points indicating higher detection rates and redder points indicating lower detection rates. Targets with the same stimulus color, background color, and illumination condition grouped closely together on the plot, indicating that luminance and contrast levels were a function of those properties. Therefore, we refer to these groups of targets with the same stimulus color, background

color, and illumination condition as “clusters.” For example, cluster 4 includes all grey targets against a black background under spotlight illumination. In Figure 2-8, clusters are marked by ovals around groups of points, along with a number to specify the cluster’s combination of target color, background color, and illumination condition, which is indicated in the table below the scatterplot. Notably, some clusters overlapped on the



White Boxes

Background	Illumination	
	Ambient	Enhanced
White	7	1
Black	3	2

Grey Boxes

Background	Illumination	
	Ambient	Enhanced
White	6	8
Black	5	4

Figure 2-8: Target detectability relative to their contrast and luminance. In this scatterplot, each point represents an individual target, with contrast on the x-axis and luminance on the y-axis. Each point is color coded to indicate the target’s detection rate across all participants. Ovals mark clusters of points that represent targets with the same combination of color, background, and illumination condition, and the number next to each oval indicates what that combination is, corresponding to a cell in the table below the plot. The green overlay indicates a region within which targets were detected at a high rate, while the red overlay indicates a region within which targets were detected at a low rate.

scatterplot, as targets with different colors and illumination conditions could nevertheless have similar luminance and contrast values.

A basic visual inspection of the pattern of detection rates, as indicated by the colors of the points in Figure 2-8, suggests that luminance and contrast impacted target visibility. The clusters with the highest detection rates (1 – 4, indicated with a green overlay) congregated in the upper right region of the plot, averaging 82% detection, while the clusters with the lowest detection rates (5 – 8, indicated with a red overlay) congregated in the lower left region, averaging 48% detection. A McNemar test (see McNemar, 1947 and Sheskin, 2011 for details) showed that the difference between these groups of clusters was statistically significant, $\chi^2(1) = 133.784$, $p < 0.001$. This indicated that there was a broad pattern of high visibility for targets with high luminance and contrast, and low visibility for targets with low luminance and low contrast.

Pairwise comparisons between clusters with McNemar tests were used to evaluate the pattern of detectability from Figure 2-8 in finer detail. Conclusions from these comparisons were consolidated into a set of three principles about how luminance and

Box 1

The following set of principles describes how luminance and contrast affect detectability of objects.

Principles of Luminance and Contrast

1. If luminance is moderate, detectability will be dependent on contrast. Likewise, if contrast is moderate, detectability will be dependent on luminance.
2. If luminance is high, detectability will be high, but will depend slightly on contrast.
3. If contrast is high, detectability will be high.

contrast affect detection rate. Those principles are listed in Box 1, followed by examples and statistical test results that illustrate each principle below.

Principle 1 describes how both luminance and contrast can affect detection rate, and was derived from two comparisons. The first comparison focused on clusters of targets with moderate luminance (75 – 150 cd/m^2), finding that detection rates were significantly higher for targets with higher contrast, in clusters 3 and 4 (mean detection rate = 80%), than for targets with lower contrast, in clusters 7 and 8 (mean detection rate = 55%), $\chi^2(1) = 48.445$, $p < .001$. This difference demonstrated that at moderate luminance levels, contrast was a major factor affecting visibility. The second comparison focused on clusters of targets with moderate contrast (40 – 75%), but differences in luminance. Compared to targets in cluster 1 (mean detection rate = 79%), which had higher luminance, detection rates were significantly lower for lower luminance targets in both cluster 5 (mean detection rate = 37%), $\chi^2(1) = 45.224$, $p < 0.001$, and cluster 6 (mean detection rate = 44%), $\chi^2(1) = 35.149$, $p < 0.001$. Mirroring our finding from the first comparison, these results demonstrate that for targets with moderate contrast levels, detection rate can either be high or low, depending on luminance levels.

Principle 2 states that targets with high luminance levels will likely have high detection rates, and while contrast remains important, it holds less influence. Principle two was based on a comparison between clusters 1 and 2, both of which had high luminance levels, but cluster 2 had a higher contrast level. The mean detection rate of targets in cluster 2 was significantly higher than cluster 1 by 9%, $\chi^2(1) = 4.971$, $p = .026$,

suggesting that higher contrast contributed to higher detection rates, but to a lesser extent than at moderate luminance.

Principle 3 describes how for high contrast targets, detection rates are likely to be high, regardless of luminance. This principle was based on a comparison between two clusters of high contrast (>90%) targets: cluster 2, which had a higher luminance level and a mean detection rate of 88%, and cluster 3, with lower luminance level and a mean detection rate of 91%. The difference in detection rate was not significant, $\chi^2(1) = .321$, $p = .571$, and illustrates how higher luminance does not always equate to higher detection rates.

Impact of Spotlight Illumination

Applying spotlight illumination always simultaneously increases luminance and changes contrast (increasing it substantially, increasing it only slightly, or decreasing it, as shown in Figure 2-5), which may or may not affect detection rate. In the context of the luminance-contrast axes, spotlight illumination therefore effectively shifts targets from one cluster to another. These shifts are illustrated in Figure 2-9, which also indicates whether or not spotlight illumination led to significant changes in detection rate. Trends in detection rate changes due to spotlight illumination were used to create a set of three guidelines that describe the impact of spotlight illumination on object visibility, included in Box 2 below, which is followed by statistical test results and examples that illustrate those guidelines.

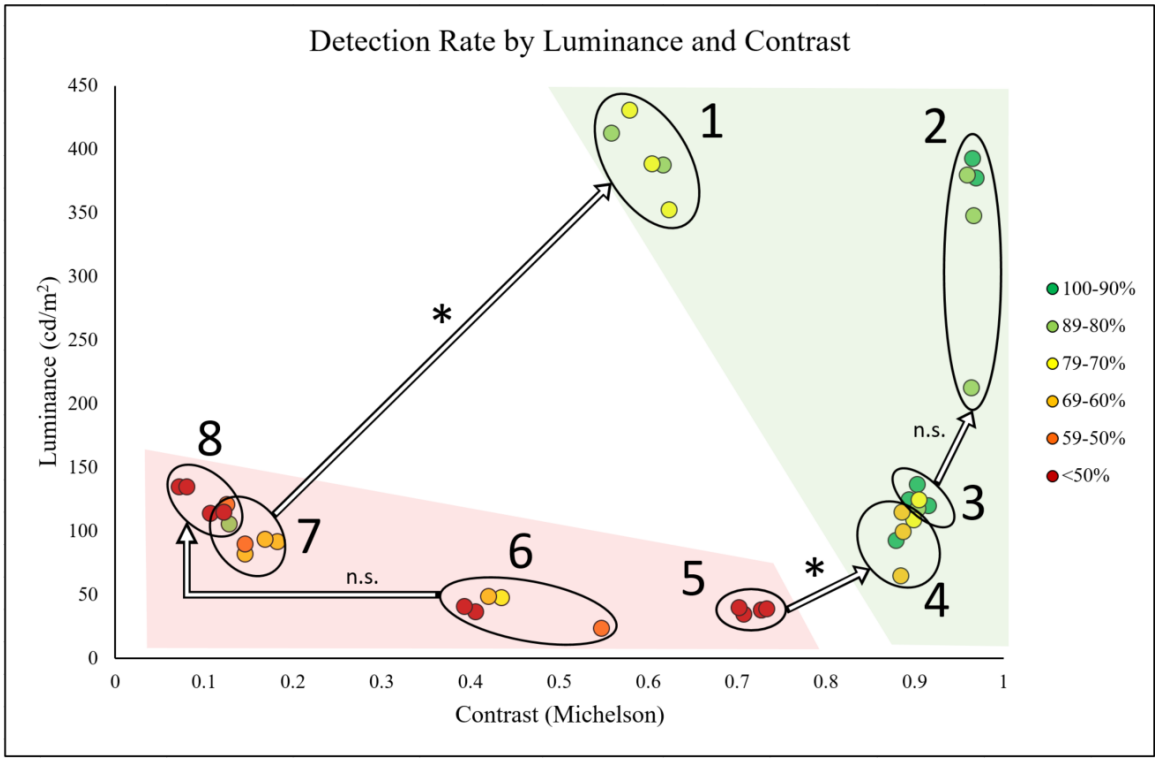


Figure 2-9: The contrast-luminance axes with the impact of spotlight illumination indicated. Black and white arrows indicate where spotlight illumination shifted targets from one cluster to another. Asterisks indicate where that shift led to significant improvements in detectability ($p < .05$), while n.s. indicates where it had no impact.

Box 2

The following is a set of guidelines that describes how spotlight illumination functions to affect object detectability, with references to the principles of luminance and contrast from Box 1.

Guidelines for Spotlight Illumination

1. If spotlight illumination is used to increase contrast, it improves detectability.
2. If spotlight illumination increases luminance but simultaneously creates low contrast, detectability will be low.
3. If spotlight illumination is used to increase luminance on a target that already has high contrast, it will have little impact on detectability, as per Principle 3.

Guideline 1 describes how spotlight illumination improved detection rates when it substantially increased contrast (in our sample, when contrast was increased by >20%).

This was the case for white targets against white walls, as spotlight illumination shifted them from cluster 7 to 1, causing high luminance and moderate contrast. This resulted in detection rates significantly increasing, from 58% to 78%, $\chi^2(1) = 15.789$, $p < .001$, as per Principle 1. Spotlight illumination also substantially increased contrast for grey targets against black doors, shifting them from cluster 5 (mean detection rate = 37%) to cluster 4 (mean detection rate = 69%), leading to a significant increase in detection rate, $\chi2(1) = 31.557$, $p < .001$. In this case, spotlight illumination improved detection rates by increased contrast to a high level, as per Principle 3.

Guideline 2 addresses a case in which increasing luminance will not improve detection rates: when spotlight illumination acts to reduce contrast. This was the case for grey targets against white walls, as spotlight illumination shifted them from cluster 6 (mean detection rate = 44%) to cluster 7 (mean detection rate = 52%), with no significant improvement in detection rates $\chi^2(1) = 1.779$, $p = .182$. This shift's effect is in line with Principle 1: even if spotlight illumination increases luminance to a moderate level, if it also causes contrast to be low, detection rate will be low.

Guideline 3 describes another case in which spotlight illumination also had little impact on detectability: when it did not substantially increase contrast. When enhanced illumination moved those targets from cluster 3 to cluster 2, while it did substantially increase luminance, there was only a marginal increase in contrast. That shift had no significant effect on detection rate, $\chi^2(1) = .321$, $p = .571$. In fact, though cluster 3 had lower luminance and contrast, it still had a higher detection rate than cluster 2, at 91% and 88% respectively. These targets may have approached a ceiling for detection rate,

with differences between clusters due to noise in performance data. This ceiling effect was likely driven by targets in cluster 3 already having high contrast, and thus high detectability, as per Principle 3.

Impact of Other Factors

While luminance and contrast drive many of the differences across clusters, they did not account for all differences in performance. For example, while clusters 3 and 4 overlap in the 2-D luminance-contrast space, detection rate was 22% higher in cluster 3, a significant difference, $\chi^2(1) = 22.765$, $p < .001$. What factors beyond luminance and contrast might account for these differences in detection rate? A major physical difference between the stimuli in these clusters is their target color – cluster 3 represents white targets (under ambient lighting, against a black background), while cluster 4 represents grey targets (under enhanced illumination against a black background). However, there is another pair of clusters of different color targets with similar luminance and contrast that do not show this difference. The white targets in cluster 7 (mean detection rate of 58%) and the grey targets in cluster 8 (mean detection rate of 52%) overlapped on the luminance-contrast axes, and their detection rates were not statistically significantly different, $\chi^2(1) = 1.422$, $p = .233$. This discrepancy suggests that in addition to luminance, contrast, and target color, some other factors were likely at play as well. A potential explanation for this discrepancy involves the viewpoints at which participants viewed the targets as they walked through the experimental space, and is explored in more detail in the Impact of Viewpoint segment of the Discussion section.

Even within clusters, where targets had the same color, background, illumination condition, and similar contrast and luminance, some targets seemed to be more visible than others. For example, in Figures 2-8 and 2-9, note the green point surrounded by dark red and orange points in cluster 7. To evaluate whether any targets were detected significantly more or less than other targets in the same cluster, a Cochran's Q test was carried out for each cluster (See Seeger et al., 1968 or Aslam, 2023 for details). Results showed that while detection rates of targets within clusters was similar for some clusters, there were significant differences within some clusters due to outliers that were detected either more or less frequently. Specifically, there were no significant differences amongst targets within clusters 1, 3, 4, and 8, but there were significant differences within cluster 2 ($\chi^2(4) = 12.364$, $p = .015$), cluster 5 ($\chi^2(4) = 10.847$, $p = .028$), cluster 6 ($\chi^2(4) = 39.212$, $p < .001$), and cluster 8 ($\chi^2(4) = 26.175$, $p < .001$).

In both clusters 2 and 8, one outlier target was detected more frequently than the others, while one outlier in cluster 5 and two outliers in cluster 6 were detected less frequently than the others. Given that the targets within each cluster all had the same color, background color, illumination condition, and similar luminance and contrast, these differences suggest that some other factor was driving these differences, which is described in Impact of Conspicuity segment of the Discussion section.

Discussion

Two experiments were conducted to evaluate the effect of spotlight illumination on the visibility of objects under conditions of reduced visual acuity and contrast

sensitivity. In Experiment 1, we explored whether some types of targets might be more difficult to see than others, and if increasing their luminance with spotlight illumination would increase their visibility. Of three hypothesized difficult types of targets, we found that only small targets placed far away from the viewer were detected less frequently. We also observed that increasing luminance via spotlight illumination improved the rate at which those targets were detected, but the same effect was not consistent across all target types; increased illumination had no significant effect on the detection rate of grey targets.

One potential explanation for why grey targets were unaffected was that while the spotlight illumination increased the targets' luminance, it also decreased their contrast with the white walls that made up the majority of the background surfaces in the testing space. To further explore how contrast contributed to target visibility, Experiment 2 more systematically controlled how spotlight illumination affected contrast. By using spotlight illumination to increase, decrease or have minimal impact on contrast, we were able to develop a set of three principles (Box 1) that describe how luminance and contrast interact to affect visibility. Applying these principles to the implementation of spotlight illumination, we developed a set of three guidelines that describe how visibility is affected by spotlight illumination (Box 2).

The principles in Box 1 and the guidelines in Box 2 are important for understanding how luminance, contrast, and spotlight illumination work under specific situations, and lead to two broad conclusions. The first is a general conclusion about lighting: both luminance and contrast are important factors to the visibility of objects for

observers with restricted acuity and contrast sensitivity. Importantly, we observed that moderate luminance (75 – 150 cd/m^2) alone was not sufficient to promote good visibility - if targets also had poor contrast, they were likely to be missed. The inverse was true as well: targets with moderate contrast (40 – 75%) and low luminance frequently went undetected. This finding aligns with prior research by Thompson et al., (2022) that has shown that the visibility of architectural features and mobility hazards is dependent upon both the luminance of those features and their contrast with their surroundings. Their team developed a system to calculate "hazard visibility scores," which combined information about the contrast, luminance, and surface properties (i.e. geometry, materials) of architectural features to quantify their visibility for low vision observers with specified visual acuity and contrast sensitivity. They found that hazard visibility scores were significantly correlated with low vision participants' ability to identify architectural features, demonstrating the score's value in predicting visibility. As feature luminance and contrast were key components of the hazard visibility score calculation, their findings and our own demonstrate that both luminance and contrast are integral to visibility.

The other key finding from this research is that directional lighting can improve the detectability of objects in the environment, but it needs to be applied strategically to create contrast in order to do so. Our data showed that when spotlight illumination substantially increases contrast, it significantly improved detection. However, in all other cases – where targets already had high contrast (i.e. white targets against a black background), or where spotlight illumination actually reduced contrast (i.e. grey targets

against a white background), the extra illumination caused no change. This finding also emphasizes the value of luminance based design, which is better suited to specifying specific objects' contrast with the surrounding environment than illuminance-based design (Thompson et al., 2022).

Impact of Viewpoint

Results from Experiment 2 showed that some clusters with similar luminance and contrast levels had similar detection rates (e.g. clusters 7 and 8), while there were significant differences between others (e.g. clusters 3 and 4). A potential explanation for that discrepancy is that as participants walked along the path, their viewpoint of a given target changed as their position relative to that target changed, bringing different contrast boundaries between the target and the background into view. When participants were directly in front of targets, the visible target-background boundaries were between the path-facing side of the target and the background, as shown in Figure 2-10, panel A. However, at other points along the path (either before or after the target), the visible boundaries were between the other sides of the target and the background, as shown in Figure 2-10, panel B. Panels C and D in Figure 2-10 show the same targets from the same viewpoints as panels A and B, respectively, but through blur foils, illustrating the impact of viewpoint with restricted vision.

In cluster 4, spotlight illumination increased contrast between the path-facing side of the targets and the background. This created a subsection of the path directly in front of the targets where the visible target-background boundaries had high contrast, and thus

high visibility. However, at other points along the path, visible target-background boundaries had lower contrast, due to other sides of the target being under ambient illumination. Therefore, while targets in cluster 4 had nominally high contrast, it was not always visible, diminishing its efficacy as a detection cue. This may have led to lower overall detection rates for targets in cluster 4 than would be expected based on their spotlight-generated luminance and contrast levels. For the white targets under ambient lighting in cluster 3, contrast was as high and viewpoint invariant, due to all sides being under ambient illumination. Thus, the visibility of targets in cluster 4 would have improved at positions closer to the targets, while the visibility of targets in cluster 3 would have remained more consistent regardless of position, which may have caused the difference in their detection rates. This concept is illustrated in Figure 2-10, Panel E.

Contrary to its function in cluster 4, spotlight illumination in cluster 8 decreased contrast between the path-facing side of the targets and the background, such that the boundaries visible from in front of the target were low contrast. From other positions along the path, contrast was moderate, but not high enough to improve detection rates. Visibility was limited from both viewpoints, and detection rates were ultimately low in cluster 8. Similarly, the targets under ambient illumination in cluster 7 also had low detection rates, which were not significantly different from those in cluster 8. This was likely due to their viewpoint invariant, low contrast boundaries with the background. Thus, a possible reason there was a difference in detection rates between clusters 3 and 4, but not clusters 7 and 8, is that changes in viewpoint as participants walked along the path had a greater impact on the visibility of targets in cluster 4 than they did for cluster

8. These viewpoint effects underscore the importance of target-background contrast in determining visibility and suggest that high contrast will be most effectively leveraged as a detection cue if it is viewpoint invariant.

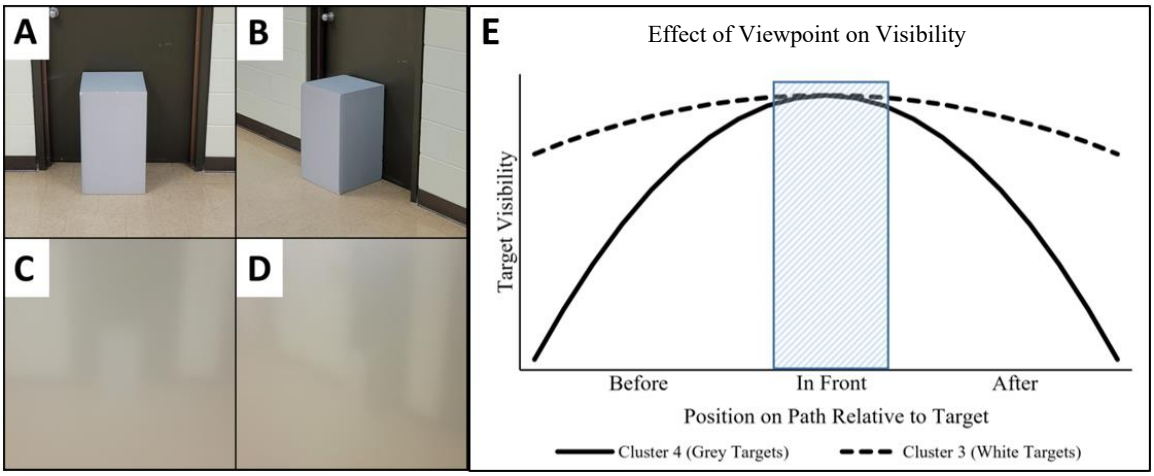


Figure 2-10: The impact of viewpoint on visibility. **A)** Photo illustrating the viewpoint on a target from a position on the path directly in front of it. **B)** Photo illustrating the viewpoint on a target from a position on the path before reaching the target. **C, D)** Photos similar in perspective to A and B, but taken through the blur foils. **E:** The conceptual mapping (not representing actual data) of how a target's visibility under spotlight illumination may have been affected by viewpoint. The x-axis corresponds to one's position on the walking path relative to a target, with the blue box representing the position directly in front of the target. The y-axis corresponds to how visible that target would be on a general spectrum from least visible to most visible. While both targets become more visible as one approaches them, due to subtending more degrees of visual angle at a closer distance, visibility is improved more for targets from cluster 4 than for targets in cluster 3.

Impact of Conspicuity

Luminance and contrast go a long way toward explaining target detection rates, but as there were significant outliers in detection rates even within clusters, there are clearly other factors at work. One plausible additional factor is visual conspicuity, which is "the degree to which an object is visually embedded in its surrounding scene" (Wertheim, 2010), or how strongly something visually stands out from the scene surrounding it. Conspicuity is determined by how an object relates to the particular

environment it is in, rather than being a property of an object itself. A major factor affecting the object-environment relationship is the density of other similar objects immediately surrounding it, with conspicuity being greater for objects in isolation. For instance, a person standing in an empty room, wearing maroon and gold overalls and waving a flag would be a highly conspicuous, easily detectable stimulus. However, that same person, when surrounded by other people wearing similar clothes, and situated in the stands at a football game would be much harder to spot. Nothing changed about the person's appearance, but surrounded by others with similar appearances, they are less conspicuous, and thus less detectable.

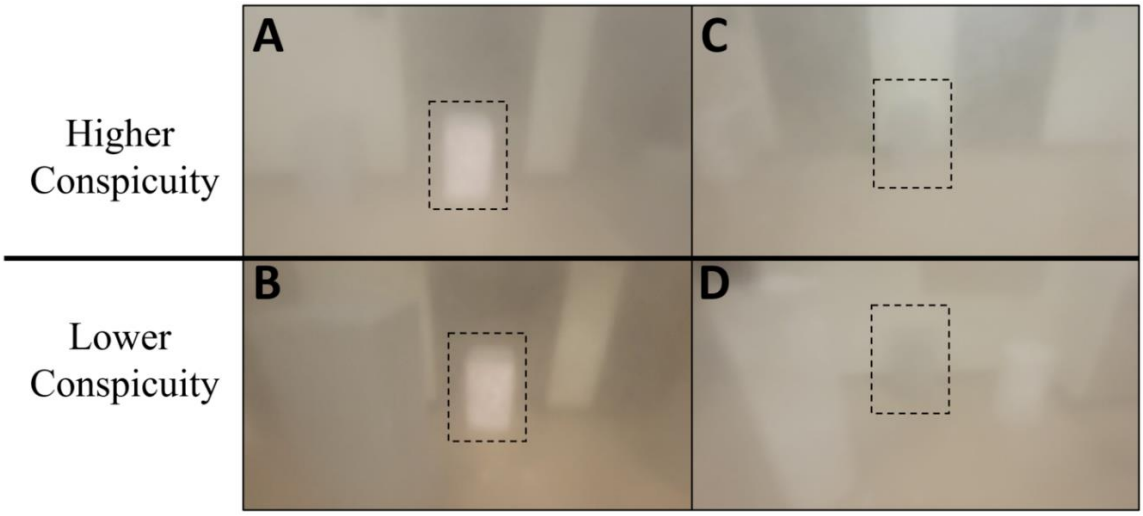


Figure 2-11: Photos of targets with especially high or low detection rates, as seen through the blur goggles, with dotted lines added around the targets in question. Photos A and C show targets that were not surrounded by other targets, potentially increasing their conspicuity. Both had significantly higher detection rates than other targets in the same cluster. Photos B and D show targets that were surrounded by other targets, potentially reducing their conspicuity. Both had significantly lower detection rates than other targets in their cluster.

Reviewing the positioning of targets that were detected significantly less frequently than others in their cluster, we observed that they were placed nearby other

targets, with distances between them less than approximately 1.5 ft. Inversely, those that were detected significantly more frequently tended to be more isolated from other targets. Examples of both are shown in Figure 2-11. The spatial layout of those targets may have affected their conspicuity, which in turn could have driven the within-cluster differences in detection rates.

Limitations and Future Research

In the current study, our participants were normally sighted, meaning our results are based on data from simulated low vision rather than naturally occurring low vision. While simulated low vision is valuable for controlling experimental parameters, it is important to recognize that it does not replicate the experience or strategies that people with real low vision will develop to aid in their mobility. For instance, Bochsler et al. (2013) reported differences in the impact of lighting and background contrast on recognizing ramps and steps between simulated low vision and real low vision.

In addition to simulated visual impairment, the environment in which we conducted these experiments was confined to a laboratory setting. Some evidence suggests that spatial learning with low vision is more difficult in more complex and irregular environments (Barhorst-Cates et al., 2019), and while our task did not involve spatial learning or memory, deployment of attention and subsequent search strategies may be different in a more naturalistic environment. Therefore, to ensure that our findings apply to environments such as homes or public spaces, it would be valuable to

conduct additional research on the impacts of directional lighting in those specific environments.

Additionally, while the precise magnitude of the impact of conspicuity and participant perspective on the detectability of targets is not clear, our results highlight how the complexity of physical environments can affect object visibility. Ecologically valid settings are particularly challenging, as the same features that make them realistic can affect the impact of other visibility factors. For instance, in a realistic café environment filled with multiple tables and chairs, a lighting enhancement might be applied to a table and not benefit its visibility – not due to limitations in the table's luminance or contrast, but because of its poor conspicuity amongst other similar objects. Therefore, a valuable direction for future research will be to clearly characterize and quantify effects arising from the configuration of objects in the environment and changes in perspective so that their impact on visibility, in conjunction with luminance and contrast, can be accounted for.

Conclusions

Our findings about the importance of contrast and its interaction with luminance have important implications for designers aiming to create visually accessible spaces. It will be important for those implementing directional lighting for the purpose of meeting luminance benchmarks to be cognizant of the interaction between focused illumination and contrast. To conclude, we found that using directional lighting to illuminate features of the environment can help people with low vision to detect them if it is used to

strategically create high-contrast cues that help differentiate the feature from the surface behind it. When working to improve the accessibility of interior designs, lighting is a powerful tool, but it must be implemented with consideration of the broader environment.

Chapter 3: Development and Validation of Online Vision Tests in Normally Sighted Participants with Artificial Acuity Reduction

Introduction

Tracking changes in visual acuity (VA), contrast sensitivity (CS), and visual field (VF) can be a useful tool for ophthalmologists to monitor the progression of eye disease in low vision (LV). Typically, these measurements of vision status are conducted in clinics, requiring patients to be present at the clinic in person. However, this requirement can be a barrier against receiving regular checkups for some patients. For instance, a study conducted by Woodward et al. in 2017 found that lack of transportation availability, work conflicts, illness, and other responsibilities such as dependent care all contributed to missed ophthalmology appointments, especially for the uninsured. More recently, a study by Brant et al. in 2021 observed an increase in missed ophthalmology appointments by more than 20% during the COVID-19 pandemic.

While the pandemic made in-person clinic visits difficult or impossible for many, it also highlighted an important potential alternative: telehealth checkups. For example, a 2021 study by Portney et al. compared the number of ophthalmology telehealth visits to in-person visits before and during the pandemic. They found that the number of appointments that were done via telehealth increased by over 20 fold, demonstrating that remote testing can be an effective tool to connect doctors and patients when in-person testing is not an option. Similarly, remote vision testing could be of value in LV rehabilitation. The prescription of magnifiers for reading is an important part of LV

rehabilitation, and recent evidence has shown that telehealth can be an effective avenue for ensuring patients are using their devices properly (Bittner et al., 2024). However, a necessary step in determining which magnifiers are most effective is measurement of VA, meaning rehabilitation specialists would benefit from access to reliable remote VA measurement. Outside of clinical and rehabilitation settings, remote testing could also be a valuable asset for researchers who conduct scientific studies of LV, where VA and CS measurement are common (Brussee et al., 2017; Mena-Guevara et al., 2021; Swenor et al., 2015; Xiong et al., 2020). Particularly for studies that can be conducted online, being able to measure VA and CS remotely could reduce expenses and save time in the experimental process.

Challenges and Limitations to Remote Testing

Remote tests must overcome several challenges related to the calibration of VA and CS testing. Calibration of VA tests depends on ensuring the test stimuli subtend the correct visual angle, which is affected by viewing distance and screen size. Establishing viewing distance in remote testing is a challenge, as it must be precise, straightforward for test takers to adjust, and cannot rely on any tools that the test taker would not have available at home. While prior studies validating remote tests have typically established precise viewing distances, most have accomplished this using experimenter assistance to hold the testing device (Bastawrous et al., 2015; Han et al., 2019; Tiraset et al., 2021; Varadaraj et al., 2021; Katibeh et al., 2022; Kim et al., 2022), who would not be there to assist in typical real testing conditions. One novel approach was to use “automatic distance calibration” function of iOS devices, which uses the devices’ front facing

camera to measure viewing distance (Wu et al., 2024), though this method excludes users who do not own Apple devices. Another method to circumvent this issue was to mail out a pre-measured string that participants could use to measure their viewing distance (Crossland et al., 2021), but the extra step and time investment associated with snail mail may be problematic for some potential users. Alongside controlling viewing distance, controlling the size of the images presented is another technical challenge for remote VA testing. Screen size can affect the size of images displayed on them, and many tests are validated only on one model of a specific device, such as a mobile phone or iPad (Habtamu et al., 2019; Varadaraj et al., 2021; Aslam et al., 2016), without clearly demonstrating that they can be used with the range of hardware available to the test taker.

Some remote VA tests have overcome the challenges with calibration and have been well validated by demonstrating good agreement with clinical measures of VA, including the Peek Acuity (Bastawrous et al., 2015), the Peek Near Vision (Katibeh et al., 2022), the V@home test (Han et al., 2019), and the Home Acuity test (Crossland et al., 2021). However, as of now, the range of those tests is limited to the better end of the VA spectrum, with the lowest VA measurable by well validated remote tests being 1.3 logMAR (Snellen equivalent of 20/400). Extending that range to poorer values of VA could be helpful for monitoring the progression of ocular disease. Even severely impaired vision with VA poorer than 1.9 logMAR (Snellen equivalent of 20/1600) is useful in day to day life (Adeyemo et al., 2017), and is thus worth monitoring to ensure timely interventions if necessary. Therefore, a test capable of evaluating severely impaired VA poorer than 1.3 logMAR would be a valuable asset in the telehealth toolkit.

CS tests must be calibrated to ensure the same contrast levels are displayed each time the test is administered. To do so, digital tests must account for the display's gamma curve, which is the relationship between pixel luminance and an image's encoded grey levels (Smith, 1995). This is a significant challenge for tests being displayed on hardware with unknown gamma curve settings, as is the case for remote tests on a patient's own device, because there can be significant variation in default gamma settings across consumer grade displays (Ghodrati et al., 2015). Unfortunately, most studies that develop remote or virtual CS tests do not implement any method to address inconsistencies in gamma curves for their tests (Hegde et al., 2023; Kollbaum et al., 2014; Rodríguez-Vallejo et al., 2014; Varadaraj et al., 2021; Kingsnorth et al., 2016; Pang et al., 2021; Kim et al., 2022; Crossland et al., 2023; Roche et al., 2023). Some discuss its importance and measure it on the specific display device used for their study, but offer little explanation of how variation across remote patients' devices might affect results (Habtamu et al., 2019; Pang et al., 2021; Varadaraj et al., 2021; Dorr et al., 2013). Others mention gamma calibration as an unresolved issue in remote testing (Richman et al., 2015; Faria et al., 2015), suggesting that it will only be an issue at low contrast levels.

Aims of Current Work

The primary aim of the current work was to develop and validate remote tests of VA and CS that extend the range of low vision that can be tested and that are practical to use across a range of typical computer displays at home. We propose that the new tests must meet the following four goals: 1) the tests need to produce results that are in good agreement with chart-based gold standard clinical measures. 2) The tests need to be easily

administered online, without the need for any special equipment, apart from a computer or iPad from the test takers. 3) The VA test must be capable of measuring severely impaired vision, well past 1.3 logMAR. 4) The tests should be quick and easy to administer, and results should be straightforward to interpret. As part of meeting goal #4, we integrated all of our tests into one software package with the aim of making testing multiple facets of vision easier and more efficient.

We were also interested in exploring whether the contrast polarity of our tests would impact either VA or CS results, as some individuals with LV find reading text with reverse contrast polarity easier (Legge, 2016). For individuals with severely impaired vision, who may not be able to recognize large optotypes under standard polarity, having a reverse polarity option may make a VA test usable. Therefore, we included both standard (black-on-white) and reverse polarity (white-on-black) tests for VA and CS.

We also developed a test of visual field (VF) size, but due to the additional complexity involved with measuring VF, validation of that test was postponed for a future study (details of the design and administration of the VF test are included in Appendix 3).

This project is described in two chapters, which represent two stages of development. This chapter is the first, and focuses on the development and technical details of the test, as well as an initial validation study conducted with normally sighted participants that took the test while using artificial vision restriction in a controlled laboratory setting. Once the tests were validated with the normally sighted group, we

moved forward to validation with participants with naturally occurring low vision, testing them remotely at their homes. The results of that testing are presented in Chapter 4.

Methods

Participants

Participants were normally sighted adults who were recruited either via word of mouth or via the Research Experience Program, a platform hosted by the University of Minnesota for recruiting undergraduate students as participants in psychology research studies. We recruited 29 participants, all of whom completed participation. Participant mean age was 26 years old, ranging from 18 - 74, with 11 being male and 17 being female. Participants were compensated either monetarily, at the rate of $15 per hour, or with course credit. Informed consent was obtained from all participants. The only inclusion criteria were having normal or corrected to normal vision and being aged 18 or over. The research methods, recruitment materials, and informed consent forms for this study were approved by the Institutional Review Board of the University of Minnesota, and all research methods adhered to the tenets of the Declaration of Helsinki.

Materials, Test Setup, and Calibration

The online vision test (OVT) is designed to be accessible to anyone with an internet connection and a computer or iPad that can join an online video conference call (the OVT can be administered over both Windows and Apple OS computers), meaning that both the administrator's computer and the test-taker's device must be capable of video calls. The administrator computer used in this study was a 14-inch Dell Latitude

laptop, while participants were asked to bring and use their own laptops or tablets for testing. If participants brought their own device, no adjustments were made to their display settings, and they were used as-is. If they did not have a device of their own, as was the case for 5 participants, a 14-inch HP Pavillion laptop was provided for them to use.

Two pieces of software were used on the administrator's computer: one to run the OVT, and one to host the video call between the administrator and the participant. The testing software was PsychoPy, a free to use psychological testing platform designed for programing and running experiments using the Python coding language (Peirce et al., 2019). The video call software used was Zoom Virtual Meetings, chosen for two primary reasons. First, while the host must have Zoom installed to initiate a video conference call, it is possible for others to join the call even if they do not have Zoom installed (calls can be joined via web browser without requiring any software installation). Second, Zoom video calling includes screen sharing functionality, which was how the OVT was transmitted to the test-taker's screen.

An overview of the test setup process is as follows: the administrator first initiates a video call, which the participant joins (either via a previously emailed hyperlink or by inputting a meeting ID on the Zoom website in a web browser). Participants are then asked ensure they do not have any magnification or color correction software active, and set their Zoom window to full screen. Next, the administrator runs PsychoPy on the administrator computer, then they initiate a screen share to display the OVT on the participant's screen. After the OVT is displayed, the participant can verbally respond to

test stimuli they see on their screen, and the administrator inputs their responses on the administrator computer.

During all testing, participants wore a set of blur goggles that restricted their visual acuity and contrast sensitivity. The blur goggles were a set of welder's goggles designed such that different lenses could be inserted in front of each eye individually. The frame of the blur goggles was made entirely of opaque plastic, such that the only light that entered the goggles was through the lenses. During testing, participants looked through a lens that was covered in Bangerter Occlusion Foils (Odell et al., 2008), while the other eye was always covered with an opaque lens blank, making viewing through the goggles monocular. The lens participants looked through did not have any refractive power, and aside from the occlusion foils, was transparent clear plastic. If a participant had prescription glasses, they wore them under the blur goggles, which had enough space to fit glasses without causing discomfort. Two types of occlusion foils were used: a Mild restriction, which reduced VA and CS to the level of moderate low vision, and a Severe restriction, which restricted VA and CS more than the foils. The specific levels of restriction for each type are described in the Results section.

Before any testing began, participants were given an outline of the testing procedure, including which tests they would be taking and what the tests were designed to measure. This outline informed participants that they would be undergoing two sets of tests designed to assess visual functioning. The first set was the newly developed Online Vision Test (OVT), which includes tests of VA and CS. They were informed that while the OVT can be administered online, their testing would be done in person, in our lab,

using their computer. The second set of tests were two widely used clinical chart-based tests, the ETDRS Visual Acuity Chart and the Pelli-Robson Contrast Sensitivity chart, and participants were given a brief description of how those tests would be administered. They were also introduced to the blur goggles, and shown the different blur foils that would be used in them. Participants were then asked to use their computer (or the one provided to them) to connect to a Zoom conference call hosted by the experimenter, and the test was transmitted to the participant's screen via screen share. After screen share was running, the participant was asked to set the Zoom call to full screen on their computer, and to turn off both their microphone and camera. This ensured that the test images were the only images being displayed on the participant's screen. After the test was set up this way, it was then calibrated.

Test calibration involved two steps, the first of which was determining how the participant's display size affected the size of test stimuli. A single white square was presented on screen against a black background. Participants, without wearing the blur goggles, were then asked to use a ruler (provided by the experimenter) to measure both the height and width of the square. This allowed for the calculation of image scaling on the participant's display by comparing the size of the squares across the two screens. For example, the square was known to be 6 cm x 6 cm on the experimenter's screen, so if the square was 8 cm x 8 cm on the participant's screen, images were 1.33 times larger on the participant's screen. This scaling factor allows for the calculation of the size of optotypes and images on the participant's screen, which was necessary for calculating the results of the VA test. The scaling factor was also necessary for ensuring the optotypes in the CS

test were size matched to the PR chart, because oversized optotypes can lead to better CS scores (Njeru et al., 2021).

The second step in the calibration process was establishing the correct viewing distance. All three of the OVT tests were taken at 12 inches viewing distance, and after calibration was complete, participants were shown how to position themselves at that distance. Participants were asked to place a 12-inch ruler (the same one used for measuring the calibration square) against their chin, and to slowly move forward until the other end of the ruler touched the center of their computer screen. After it touched the screen, they were told to remove the ruler, and remain at that distance from the screen for the duration of all tests. While the ruler was used in this study, the OVT can be calibrated for viewing at other distances, and other common household objects can be used to establish those distances. For example, paper towel rolls are standardized to the length of 11 inches in the United States, and can be used in the same fashion as the ruler to establish viewing distance. Additionally, prior research has shown that participants in online visual psychophysics studies were able to effectively maintain a consistent viewing distance with their screens during testing (Li et al., 2020). As participants in that prior study maintained their viewing distance for 10 minutes, and each part of the OVT took no longer than 5 minutes, we utilized this method as well. Viewing distance was re-measured after each test to both ensure viewing distance was consistent.

Before testing began, participants were taught how to respond to tumbling E optotypes via practice trials, conducted without the blur goggles. A tumbling E stimulus was displayed in standard polarity on the participant's screen, sized to fill most of the

screen. The E was displayed at one of four orientations, 0°, 90°, 180°, or 270°, chosen at random. Positioned at the 12-inch testing distance, participants verbally reported the orientation of the E, after which it was replaced by another E at one of the four orientations, again chosen at random. Participants were given feedback on their accuracy, and continued until they correctly identified all four orientations.

Tests were conducted in the same order for each participant: o-VA, o-CS, ETDRS, PR. For the o-VA and o-CS. Participants took each test with each eye individually and with both blur levels. Both the o-VA and o-CS had two conditions: standard polarity (black optotype against a white background) and reverse polarity (white optotype against a black background), and the standard polarity was tested first. This created a 2x2x2 setup for each test, with eye (left/right), blur level (mild/severe), and polarity (standard/reverse) at each level, meaning that each test was taken in eight conditions.

o-VA Tests

The o-VA is a 4 alternative forced-choice (4AFC) tumbling E style test, wherein a single E optotype is presented to the participant (with no flankers) at one of four orientations: 0°, 90°, 180°, or 270°. The test begins by displaying the largest E that can be accommodated by the participant's screen. Participants then verbally report the orientation of the E, and the administrator inputs their answer on the administrator computer using the arrow keys. After each response input, the E disappears, to be replaced by another E after a 500 ms delay. The test progresses in a four down, one up

staircasing procedure, with each step changing the size of the E by 0.1 log units. The test terminates at 5 reversals, leading to an 84.15% correct performance threshold (Kingdom & Prins, 2016). Participants are asked to report an orientation for the E for all trials, even if they need to guess.

To speed up testing, participants are asked on the first trial if they can easily determine the orientation of the E, or if it is at all difficult. If they report they can easily determine it, the size of the E is reduced by 0.3 log units, and they are asked again. This is repeated until the participant reports it has become difficult, after which the staircasing procedure commences.

The total range of acuities testable by the o-VA spans 2.3 logMAR (e.g. from 0.3 – 2.6, 0.4 – 2.7, 0.6 – 2.9, etc.), with the range depending on the size of the participant's display and their viewing distance. Larger displays and closer viewing distance shift the range toward poorer VA, and smaller displays and farther viewing distance shift the range toward better VA. Details about how to calculate shifts in the range of the o-VA based on different viewing distances and displays are included in Appendix 1. In the current study, the range on the smallest display was from 0.07– 2.37 logMAR, and the range on the largest screen was from 0.56 – 2.86 logMAR. Table 3-1 provides examples of different possible test ranges, given different display sizes and viewing distances.

	Display Size (Type of Display)		
Viewing Distance	10.2 inch (Small iPad)	14 inch (Laptop)	43 inch (very large desktop monitor)
11 inches	0.1 – 2.4 logMAR	0.3 – 2.6 logMAR	0.82 – 3.12 logMAR
12 inches	0.06 – 2.36 logMAR	0.26 – 2.56 logMAR	0.78 – 3.08 logMAR

Table 3-1: Example o-VA test ranges on different size displays at different testing distances. Viewing distances are either 11 or 12 inches, corresponding to the lengths of common household objects of standardized sizes that participants are instructed to use for measuring their viewing distance. Display sizes listed are the length of the diagonal of the screen.

As the E stimuli are composed of individual pixels on a digital display, their possible stroke widths are constrained to multiples of the width of a single pixel. This creates two gaps in the 0.1 log unit size progression at the better end of the VA spectrum. The first occurs in the transition between the smallest E with a stroke width of 1 pixel and the next smallest, with a stroke width of 2 pixels. This step doubles the stroke width of the stimulus, equivalent to an increase of 0.3 log units in size. The next gap occurs in the transition between 2 and 3 pixels, which is equivalent to an increase of 0.2 (rounded up) log units in size. Details of the size progression of the E stimuli, with corresponding and stroke width/pixel count, are included in Appendix 1.

o-CS Tests

The o-CS utilizes the same tumbling E and staircasing procedure as the o-VA (a 4AFC Tumbling E orientation discrimination task, with 4 down-1 up staircasing terminating at 5 reversals). However, instead of the E changing in size, the E changes in 0.1 log contrast steps. The contrast levels in the o-CS are defined with Michelson contrast, and ranged from 0.1 logCS to 2.0 logCS. Michelson contrast is typically used for defining contrast in square wave gratings, but was used here because the primary

features being detected in a tumbling E optotype are the arms, which are similar to a square wave grating with 2.5 cycles. However, the PR chart, which is a standard measure the o-CS is being compared with, uses Weber contrast to define its log steps. Therefore, for the purposes of comparing results between the tests, all results from the o-CS were converted to Weber contrast. Details about the conversion process are available in Appendix 2.

While the contrast between the E optotype and the background changes with the staircasing procedure, the size of the E optotype remains consistent across all trials. Before testing began (during calibration), the optotype size to display on the participant's screen is calculated such that the visual angle subtended by the E on their screen will match the visual angle subtended by optotypes on the PR chart, approximately 2.8° with the mild blur (testing done at 1 m), and 5.6° for the severe blur (testing on the PR chart was done at 0.5 m for the severe blur, as participants were unable to read letters at 1 m). Once the correct size was displayed, the testing for this study started in a similar way to the o-VA testing: participants were asked if they could clearly see the E. If they could clearly see it, the contrast was reduced in steps of 0.3 log contrast until the participants reported it was becoming more difficult, at which point the standard staircasing procedure began.

To address the differences in gamma calibrations across different screens, we utilized dither patterns. Dither patterns are patterns of white and black dots, arranged such that they appear noisy and avoid repeating patterns or structure, which can be used to create shades of grey. This technique is most commonly used in newspapers and

printing in greyscale, where dots of black ink are sprayed onto paper at different densities of dots per unit area to create the appearance of grey, when in fact the only ink being used is black. We created dither patterns out of 10x10 grids of pixels with different proportions of white and black pixels, with each white-black proportion corresponding to a different average luminance. These dither patterns were created in a series that changed in 0.1 log contrast steps.

As the black pixels emit very little light, typically fewer than 2 cd/m^2 even on bright screen settings, the illumination emitted by a screen using these dither patterns is a function of the proportion of white and black pixels on the screen. Therefore, grey levels created using these dither patterns are not affected by the specific gamma calibration of the screen they are displayed on. More details about the production of the dither patterns, including how the grids of pixels were created and empirical measurements of light emitted across multiple test displays, as well as examples of the dither patterns, are included in Appendix 2.

When the density of dots in the dither patterns exceeds the observer's acuity limit, the pattern of white and black pixels blurs together into a uniform, non-textured grey. At a 12-inch viewing distance, this occurs for observers with a VA poorer than approximately 0.3 logMAR (Snellen equivalent of 20/40) on most displays. Observers with VA of 0.3 logMAR or better could plausibly resolve the individual pixels in the dither patterns, meaning they would see the dither pattern as a textured surface instead of a grey field. Some evidence has shown that the visual system is more sensitive to boundaries between textured surfaces than to those between uniform surfaces (DiMattina

et al., 2022), so it is recommended that the o-CS be used only with patients with a VA poorer than 0.3 logMAR.

Chart-Based Testing

Standard clinical testing for VA was done using an ETDRS chart (Precision Vision, La Salle, IL, USA). The chart was uniformly illuminated using fluorescent lights positioned adjacent to the chart, as well as overhead illumination, such that the mean luminance of the white background of the chart was approximately 110 cd/m^2. Due to difficulty in reading the letters beyond the first few lines at the typical 4 m testing distance (and at an initial adjusted distance of 2 m) in pilot testing, participants were tested at a 1 m viewing distance for the mild blur, and a 0.5 m viewing distance for the severe blur. Participants were asked to start at the top line of the chart, and read each letter on a line before moving on to the next line. When participants approached their VA threshold, they were asked to continue attempting to read letters, guessing if necessary. This continued until participants were unable to correctly report any of the optotypes on a given line, at which point testing stopped. VA was calculated by starting at 1.0 logMAR at the top line of the chart, and reducing the logMAR value by 0.02 for each letter correctly identified. Additionally, to account for reduced viewing distance, 0.6 was added to the logMAR value in the mild blur condition, and 0.9 was added in the severe blur condition.

Standard clinical testing for CS was done using a PR chart. The chart was illuminated in the same way and to approximately the same luminance as the ETDRS

chart. Participants were tested at 1 m for the mild blur, and at 0.5 m for the severe blur (adjustment was made because several participants were unable to read any of the letters at 1 m with severe blur in pilot testing). They were asked to start at the beginning of the first line, and to read across each row as far as they could. Testing stopped when participants reported none of a given triplet correctly. Any correct answers in a triplet were counted toward their threshold score. CS was calculated by starting at 0 logCS and adding 0.05 for each letter correctly identified, as described in Elliott et al. (1991) and Arditi (2005).

Analysis

Initial analysis focused on determining whether the OVT was in good agreement with standard clinical chart-based tests. Agreement was assessed in two ways: first, we evaluated the magnitude of the differences between the mean scores from the OVT tests and their corresponding chart-based tests. This was done by conducting paired t-tests to determine if they were statistically significantly different from each other.

Next, we evaluated the variance of differences between the OVT and the charts by calculating the limits of agreement (LoA) between the tests. An LoA is a measure of agreement between two tests, and is calculated as the mean of the differences between two tests ± 1.96 SD of the differences, which provides upper and lower bounds between which 95% of the differences between the tests are likely to fall (Bland & Altman, 1996). The width of an LoA, or the range between its upper and lower bounds, corresponds to the amount of variance in the differences. Narrower LoAs indicating differences clustered

more tightly around the mean difference, indicating lower variance and better agreement, and wider LoAs indicating the differences were more dispersed, indicating higher variance and poorer agreement. A prerequisite to LoA analysis is normally distributed data, so the distribution of the differences between each set of tests was assessed using a QQ plot and found to be normally distributed before undergoing analysis.

Additional analysis focused on whether there would be any differences between the standard and reverse polarity versions of the tests. To address this question, we compared the mean scores of the tests across polarities using paired samples t-tests.

Results

Visual Acuity

Mean VA values for each test condition, mean differences between VA values from the o-VA and ETDRS chart, and LoAs between the o-VA and ETDRS chart are listed in Table 3-2. For both blur conditions and both contrast polarity versions of the o-VA, mean VA values from the o-VA were statistically significantly different from ETDRS values. In three out of four comparisons between the o-VA and the ETDRS chart, VA was better when measured by the o-VA. This was the case for the reverse polarity test in the mild blur condition, $t(55) = 8.264$, $p < 0.001$, and for both the standard polarity and reverse polarity tests in the severe blur condition, $t(55) = 8.27$, $p < 0.001$ and $t(55) = 18.638$, $p < 0.001$, respectively. The exception to this trend was in the standard polarity test in the mild blur condition, where mean VA was significantly worse than the ETDRS chart, $t(55) = 5.073$, $p < 0.001$. For both standard polarity tests, the differences

equated to approximately one line on the chart. For both the reverse polarity tests, these differences equated to approximately two lines on the chart.

Blur Condition	Test	Mean VA	Difference from ETDRS	LoA Range [Upper Bound, Lower Bound]
Mild	ETDRS	0.74 (±0.2)	---	---
	Standard Polarity o-VA	0.86 (±0.19)	-0.12*	±0.35 [-0.48, 0.23]
	Reverse Polarity o-VA	0.52 (±0.23)	0.22*	±0.34 [-0.17, 0.61]
Severe	ETDRS	1.85 (±0.11)	---	---
	Standard Polarity o-VA	1.74 (±0.11)	0.11*	±0.21** [-0.09, 0.32]
	Reverse Polarity o-VA	1.61 (±0.11)	0.24*	±0.20** [0.05, 0.44]

Table 3-2: Results from VA Testing. All data are presented in logMAR. A * indicates a statistically significant difference from the ETDRS chart, while a ** indicates an LoA that is within one letter of the width of test-retest values.

The LoAs between the o-VA and ETDRS chart were wider when participants were tested with mild blur compared to when they were tested with severe blur. This indicated there was more variance in the differences between the tests in the mild blur condition, which is illustrated in Figure 3-1 using four Bland-Altman plots, one for each o-VA condition. For the mild blur condition, the ±0.35 logMAR width of the LoAs indicates that differences in the tests would be within 3.5 lines of the mean difference 95% of the time. For the severe blur condition, the ±0.2 logMAR width of the LoAs indicated that differences would fall within 2 lines of the mean difference 95% of the time.

Lastly, mean VA measured on the reverse polarity version of the o-VA was statistically significantly better compared to the standard polarity version in both the mild blur, $t(55) = 13.354$, $p < 0.001$, and the severe blur, $t(55) = 12.051$, $p < 0.001$, conditions. In the mild blur condition, reverse polarity was better by 0.34 logMAR (equivalent to approximately 3.5 lines on the ETDRS chart). In the severe blur condition, reverse polarity was better by 0.13 logMAR (equivalent to approximately 1.25 lines on the ETDRS chart).

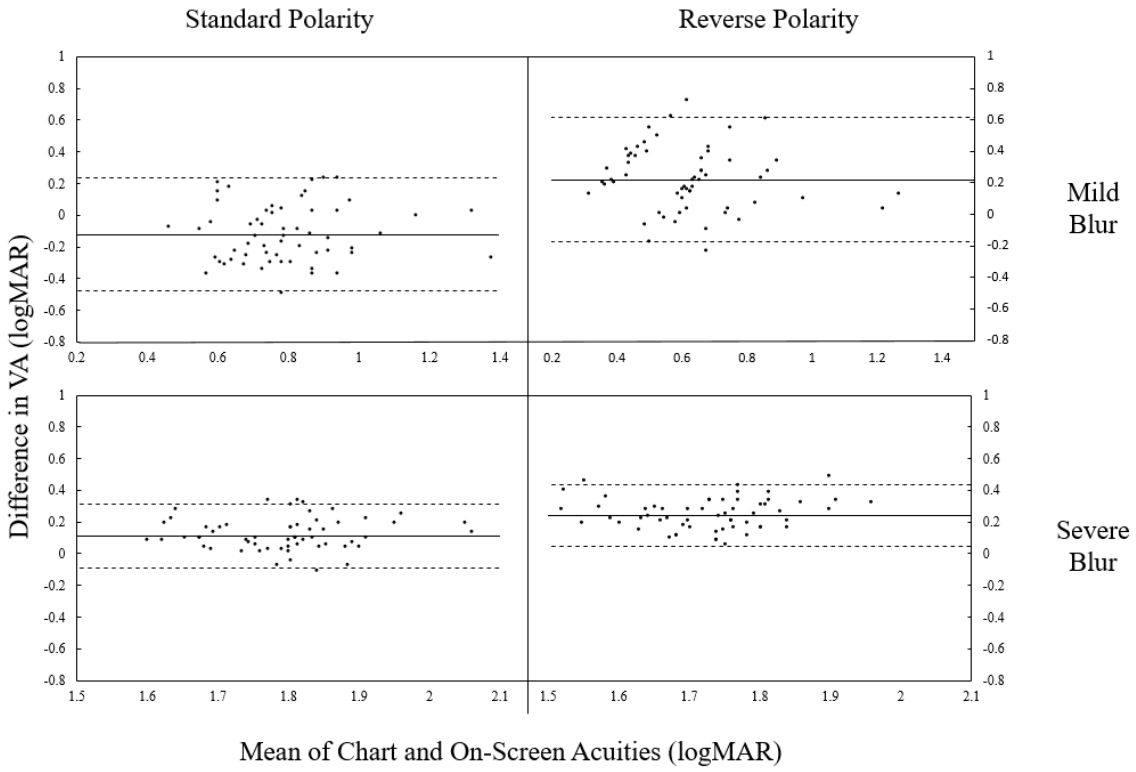


Figure 3-1: A composite of four Bland-Altman plots, one for each o-VA test condition, which illustrate the differences between the o-VA and ETDRS values. Plots are organized by polarity in the columns (left = standard polarity, right = reverse polarity) and restriction condition in rows (top = mild, bottom = severe). The x-axis of each plot is the mean VA as measured by each test, while the y-axis is the difference between the tests (ETDRS – o-VA) and each point is an individual eye. Note that the y-axes cover the same range on all four plots, while the x-axes are the same within rows. LoAs are denoted by the horizontal lines in each plot – the solid line is the mean difference, while the dashed lines are the upper and lower bounds.

Contrast Sensitivity

Mean CS values for each test condition, mean differences between CS values from the o-CS and PR chart, and LoAs between the o-CS and PR chart are listed in Table 3-3. In all conditions, mean CS scores were statistically significantly better from the o-CS than from the PR chart. In the mild blur condition, differences were small, at -0.06 logCS for standard polarity, $t(55) = 2.587$, $p = 0.012$, and -0.05 logCS for reverse polarity, $t(55) = 2.67$, $p = 0.01$, equivalent to approximately one letter on the PR chart for both. In the severe blur condition, the differences were larger, with standard polarity being different by -0.12 (equivalent to ~2 letters), $t(55) = 5.035$, $p < 0.001$, and reverse polarity being different by -0.23 logCS (equivalent to ~5 letters), $t(55) = 9.91$, $p = <0.001$.

Blur Condition	Test	Mean (logCS ±*SD*)	Difference from PR	LoA Range [Upper Bound, Lower Bound]
Mild	PR	1.09 (±0.19)	---	---
	Standard Polarity o-CS	1.15 (±0.26)	-0.06*	±0.38** [-0.44, 0.31]
	Reverse Polarity o-CS	1.14 (±.22)	-0.05*	±0.30** [-0.35, 0.25]
Severe	PR	0.49 (±0.14)	---	---
	Standard Polarity o-CS	0.61 (±0.2)	-0.12*	±0.35** [-0.46, 0.23]
	Standard Polarity o-CS	0.72 (±.0.20)	-0.23*	±0.35** [-0.57, 0.12]

Table 3-3: Results from CS Testing. All data are presented in logCS. A * indicates a statistically significant result, while a ** indicates an LoA that is within one letter of the width of test-retest values.

The widths of the LoAs between the o-CS and PR chart were consistent across all testing conditions. This meant that the variance in the differences between the o-CS and

PR chart were similar regardless of blur level or contrast polarity. In each case, the width of the LoAs indicated that differences would fall within 6-7 letters of the mean difference 95% of the time. The LoAs are visualized in Figure 3-2, which is comprised of four Bland-Altman plots, one for each o-CS condition.

Lastly, there was not a statistically significant difference between the standard and reverse polarity versions of the test in the mild blur condition, $t(55) = 0.613$, $p = 0.542$, but there was a difference between polarities in the severe blur condition, $t(55) = 5.692$, $p < 0.001$. The mean difference across polarities in the severe condition was 0.11 logCS, equivalent to approximately 2 letters on the PR chart.

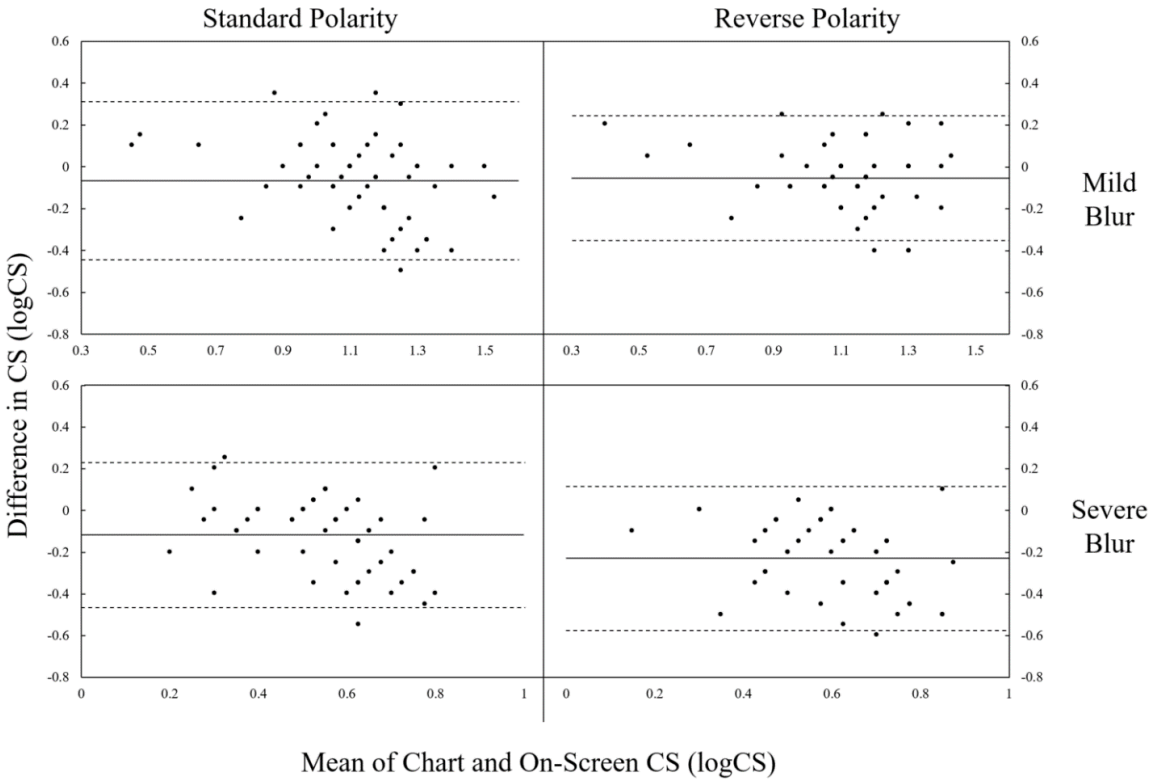


Figure 3-2: A composite of four Bland-Altman plots, one for each o-CS test condition, which use the same formatting as Figure 3-1.

Discussion

When developing the OVT, we identified four major goals that the test needed to achieve. The first goal was that the tests need to produce results that are in good agreement with gold standard clinical measures, the ETDRS and PR charts. We assessed agreement in two parts: with regard to the magnitude of the differences between the tests, using paired samples t-tests, and with regard to the variance of the differences between the tests, using LoAs.

In terms of the magnitude of the differences between the tests, we found that while there were statistically significant differences between the mean scores of the OVT and the chart-based tests, the differences were fairly small. For both the o-VA and the o-CS, in both blur conditions, the mean differences with the charts was no greater than 0.12 log units. The magnitude of these differences with the chart-based tests is similar to the differences observed in other online tests of VA and CS (Samanta et al., 2020; Mena-Guevara et al., 2021). The differences were greater in reverse polarity, but were no more than 0.24 log units. In all but one testing condition (the mild blur, standard polarity o-VA), VA and CS were better on the OVT compared to the charts. Given that the 4AFC task utilized in the OVT is simpler than the letter recognition task used in the charts, these differences are not surprising, as simpler tasks have been found to result in better performance on vision tests (Kitchen & Bailey, 1981; Bailey et al., 2012; Chaiktmongkol et al., 2018; Treacy et al., 2015).

The amount of variance in the differences between the tests is a key part of evaluating agreement because even if the results of the tests are typically different, if they are different by a consistent amount, the results of one test can still be used to predict the results of the other. However, there is no universal threshold at which the variance of the differences between tests can be considered to be "acceptably low," so it is necessary to evaluate LoAs in the context in which the tests are used. For the OVT and clinical charts, we chose the test-retest reliability of the charts to serve as a benchmark for how much variance is expected in gold-standard measurement tools. If the range between the bounds of the LoA is as small as or smaller than the test-retest reliability of the charts, we propose the variance reaches the level of industry-standard tools, and is a demonstration of very good agreement. The ETDRS chart has been shown to have a test-retest reliability of ±0.2 logMAR, and the PR chart has been shown to have a test-retest reliability of ±0.35 logCS in tests with LV participants (Kiser et al., 2005). Therefore, we consider the tests to have demonstrated very good agreement, with regards to the variance of the differences between the tests, if the width of the LoAs is equal to or less than ±0.2 logMAR for the o-VA and ±0.35 logCS for the o-CS. This method for evaluating agreement between tests by comparing test-retest reliability to LoAs is consistent with other studies validating novel vision tests (Pang et al., 2021; Kollbaum et al., 2014; Habtamu et al., 2019). We found that the variance in the differences between the tests met those thresholds for the o-VA and ETDRS chart under severe blur in both contrast polarities, and for the o-CS and PR chart in all testing conditions. When viewed with mild blur, there was more variance in the differences between the o-VA and ETDRS

chart, such that the differences were less reliable than the test-retest reliability of the chart.

A possible explanation for the increased variance in the difference scores is that near the better end of the VA spectrum, the 0.1 log size unit progression of the o-VA stimuli is interrupted, as explained in the o-VA segment of the Methods section. Steps are greater than 0.1 log size unit between the smallest 3 stimuli, reducing the test's accuracy at that range. This issue was most prevalent for the reverse contrast polarity version of the test, where 27 of our 28 participants reached one of the smallest three stimuli, whereas only 18 participants reached that level in the standard polarity version. This suggests that the o-VA is best suited for measuring moderately or severely reduced VA, compared to near normal VA, particularly in the reverse contrast polarity version.

Evaluating agreement in terms of both the magnitude and variance of the differences between the tests, we categorized each combination of test, contrast polarity condition, and blur level as having very good, good, poor, or very poor agreement. Combinations with small differences that had low variance had very good agreement, with typically tightly aligned scores. Combinations with moderate differences, but low variance in those differences, had good agreement, as OVT results could be used to predict chart results both at a group mean level and at an individual level by accounting for mean differences. Combinations with small differences but high variance had poor agreement, because OVT scores were predictive of chart scores at the group mean level, but not at an individual level. Those with large differences that were also highly variable had very poor agreement, as scores did not align at either at a group or an individual

	Low Differences Variance	**High Differences Variance**
Small Differences	Very Good Agreement • o-VA [standard, severe] • o-CS [standard, mild] • o-CS [reverse, mild] • o-CS [standard, severe]	Poor Agreement • o-VA [standard, mild]
Moderate Differences	Good Agreement • o-VA [reverse, severe] • o-CS [reverse, severe]	Very Poor Agreement • o-VA [reverse, mild]

Table 3-4: Categorization of different combinations of OVT test, contrast polarity version, and blur condition based on their level of agreement with the corresponding chart. Each bullet point in a cell corresponds to a single combination, listing first which test it is, followed by brackets that specify which contrast polarity condition it is (standard or reverse), followed by the blur condition (mild or severe).

level. Each combination of test, contrast polarity, and blur condition is categorized in this way in Table 3-4.

Notably, only the standard polarity o-VA with mild blur was categorized as having poor agreement, and only the reverse polarity o-VA with mild blur was categorized as having very poor agreement, with all other combinations demonstrating either good or very good agreement with the charts. With the majority of testing conditions demonstrating either good or very good agreement with the charts, the OVT met our first goal of being a remote test that is high in agreement with the standard clinical chart-based tests.

The second goal was that the tests need to be effectively administered online without requiring any special equipment, aside from a computer or iPad that can connect to an online video call. The OVT met this goal, and utilizes common household items (in this case, a standard 12-inch ruler) for calibration. While viewing distance stability was

not recorded, our “common household items” method shows promise because it was effective in producing tests that were typically in good or very good agreement with gold-standard charts, which would not be possible with unstable viewing distances. Furthermore, no participants reported fatigue during testing or requested breaks from testing, suggesting they were able to comfortably maintain their viewing distance without the need for a chinrest or other stabilizing equipment.

The third goal was that the tests need to be capable of measuring severely impaired VA, past 20/400, which is the poorest VA measurable by other remote tests. Even on a small display, such as an iPad with a 10.2-inch screen, and at an 11-inch viewing distance, the OVT can measure a VA of 2.4 logMAR, or approximately 20/5000, meeting the goal of measuring severely impaired VA.

Given this range, the OVT is capable of measuring VA into the range of “ultra-low vision,” a category of sight characterized by being unable to resolve even the largest letters on the ETDRS chart at a 0.5 m testing distance (quantitatively, VA poorer than 20/1600) (Geruschat et al., 2012). Due to being outside the range of the ETDRS chart, VA at this level is typically qualitatively categorized in clinical settings (e.g. “finger counting” or “hand motion) rather than quantitatively measured. Some specialized tests do exist for the quantification of ultra-low vision, including the Berkeley Rudimentary Vision Test (Bailey et al., 2012) and the Freiburg Acuity and Contrast Test (Bach, 1996, 2006). However, the Berkeley test cannot be administered remotely, and while the Freiburg test is computerized, it has not been validated in remote testing. Filling that gap, the o-VA shows promise in its potential to remotely quantify VA in ultra-low vision.

The fourth goal was that the tests should be simple and easy to administer, and results should be straightforward to interpret. Administering the o-VA is as easy as hosting a Zoom call, starting a screen share, and inputting participant responses with the arrow keys. The o-CS is similarly straightforward, with the addition of setting the E stimuli to the appropriate size, as explained in o-CS segment of the Methods section. Notably, in some cases, severe restrictions in VA may prevent some test takers from resolving the E at that correct size. In such cases, if both VA and CS are to be tested, it is advisable to first test VA to determine which size E they can reliably resolve. Once determined, the best approach is to present an E well above their VA threshold for the o-CS, to ensure than any errors are due to low contrast, rather than to limitations due to VA. However, if this is done, it is important to note that this will likely result in a CS better than what would be measured on a PR chart at the standard 1 m testing distance.

It is also worth addressing the experience of running the OVT from the perspective of a test taker. While data was not collected on participant comfort level, they were frequently asked whether or not they were experiencing any neck or eye strain, and whether they wished to take breaks. Requests for breaks were rare (only occurring once or twice), and there were no reports of eye strain. Furthermore, while data was not gathered on the administration time for any of the tests, administration of the OVT tended to go quickly, with each individual test taking no more than approximately 3 minutes.

Contrast Polarity Effect

In addition to our four primary goals, we were also interested in exploring whether the contrast polarity of our tests would impact either VA or CS results, as some individuals with LV show improved reading performance with reverse contrast polarity text (Legge et al., 1985; Ehrlich, 1987; Wu et al., 2020). This benefit is prominent in diseases that cause cloudy ocular media, as the cloudy media can scatter light within the eye, which effectively reduces perceived contrast. As standard polarity text emits more light than reverse polarity text, the scattering effect is stronger, leading to worse reading performance. Here, we found improved performance in reverse contrast polarity for the o-VA in both mild and severe restriction, and for the o-CS under severe restriction. Contrast polarity had the largest impact on the o-VA under mild blur, where the mean scores were 0.34 logMAR better (equivalent to approximately 3.5 lines on the ETDRS chart) in reverse compared to standard polarity. Additionally, the standard polarity o-VA under mild blur was the only condition in which mean scores on an OVT version of a test were worse than their chart counterpart. This may have occurred because that was the condition in which the dark E stimulus was smallest, leaving the majority of the screen to display the white background, thus maximizing the light scattering effect of the goggles. These effects suggest both that our blur goggles likely mimicked the light scattering effect of cloudy ocular media, and that the OVT is sensitive to contrast polarity effects.

Limitations

There are some limitations to the OVT that should be considered. As discussed in the o-VA and o-CS segments of the Methods section, the design of the stimuli used in both the o-VA and o-CS puts constraints on testing at the better end of the VA spectrum. For the o-VA, the test is typically constrained to testing LV with a VA poorer than 20/40. A plausible solution for this problem is increasing viewing distance, which reduces the visual angle subtended by each pixel, causing the stimuli to require better VA to resolve. However, that increased distance would require more careful measurement on the part of the test taker, and that process was not evaluated here. The 0.1 log unit step size progression is also constrained at better VA as well, reducing accuracy for near normal vision.

For the o-CS, viewers with sufficiently high VA to resolve individual pixels could affect measurement by changing the viewers' perception of the grey stimuli from a smooth, uniform surface to a textured one comprised of white and black dots. As sensitivity to differences between uniform and textured surfaces may be different (DiMattina et al., 2022), this could affect the CS results for those with better VA.

Future Directions

Given the generally good performance demonstrated by the OVT for normally sighted subjects with artificial acuity reduction in this preliminary group, our next research step will be to continue validating the test by evaluating it in a study with participants with naturally occurring low vision. Additionally, we note that the current

study was conducted in a laboratory setting, with participants who were primarily young undergraduate students. Future research will need to determine whether at-home testing will affect the efficacy of the test. Also, as the majority of individuals with low vision are of older age, it will be important to determine if familiarity with technology, which is typically high in young people, but may be lower in an older population, will have any effect on how easy it is to administer the test.

Conclusion

In conclusion, our findings here support that the OVT is effective at measuring both VA and CS online. The results from the OVT are in good agreement with gold standard, chart-based clinical tests of both VA and CS. Both tests can be administered effectively online, on both Windows and Apple displays, across a broad range of display sizes, regardless of the gamma correction calibration of the displays. The tests were also simple and easy to administer, taking only a few minutes per test, and the results are on the same scales used by their standard clinical chart counterparts, making interpretation of results intuitive. Lastly, the OVT extends the range of quantifiable VA testing into ultra-low vision, and future research will be conducted to determine how effective the test is in participants with that level of vision.

Chapter 4: Validation of Online Tests of Visual Acuity and Contrast Sensitivity in Low Vision Participants

Introduction

Low vision (LV), defined by the International Classification of Diseases and Related Health Problems as having visual acuity poorer than 20/40 (Crossland et al., 2021), is a prominent disability in United States with nearly 6 million Americans living with vision loss as of 2017. Furthermore, that figure is expected to rise to nearly 10 million in the next 30 years (Chan et al., 2018). Worldwide, the trend is similar, with 237 million people living with vision loss in 2020, and current predictions show that number is likely to rise to 588 million by 2050 (Bourne et al., 2017). Impairments in vision can reduce one's ability to carry out tasks of daily living, and can have a substantial negative impact on one's quality of life (Varadaraj et al., 2021). For those with LV, regularly monitoring vision status is an important part of effectively managing vision loss and mitigating progressive impairment.

Typically, regular vision checkups involve measurement of visual acuity (VA), which characterizes the ability to discern fine details in scenes. Reductions in VA can be warning signs for the progression of ocular disease, so regular checkups are an important part of treating those conditions (Ho et al., 2020). While VA does not necessarily reliably predict performance on tasks of daily living (Legge et al., 1985), it is still an intuitive and informative metric that most patients understand. Measuring VA is also an important part

of LV rehabilitation, as it is important for the prescription of optical magnifiers for reading (Bitner et al., 2024), one of the most frequently utilized rehabilitation services.

Another important facet of visual function is contrast sensitivity (CS), or the ability to differentiate between shades of grey. This fundamental property of vision is integral to the processes of differentiating objects from their surroundings (Chung & Legge, 2016), maintaining stable and balanced mobility (Swenor et al., 2015), and reading speed (West et al., 2002), and higher contrast sensitivity is positively correlated with measures of vision related quality of life (Eshraghi et al., 2019). CS is also useful for monitoring the progression of ocular disease, including glaucoma (Onoja et al., 2023) and several maculopathies (Wai et al., 2022).

Both VA and CS loss can occur simultaneously in eye disease, and several studies have documented correlations between the two across a variety of pathologies, including age-related macular degeneration (Bellman et al., 2003), cataract (Elliott & Situ, 1998), and retinitis pigmentosa (Alexander et al., 1995; Kiser et al., 2005). However, recent evidence has shown that while there may be simultaneous loss of VA and CS in LV, the degree of the loss in each dimension can differ. Xiong et al. (2020) evaluated both VA and CS in groups of LV participants with several ocular pathologies. They found differences in the quantitative relationship between VA loss and CS loss across diseases, suggesting that a thorough evaluation of vision status should include measurement of both VA and CS.

Unfortunately, regular visits to ophthalmology clinics for vision checkups can present a significant challenge. A study conducted by Woodward et al. in 2017 found that many patients at ophthalmology clinics miss appointments due to a lack of transportation availability, illness, and dependent care, all of which can make a trip to the clinic difficult. Another study surveyed 764 patients from low vision clinics around the United States, finding that more than half reported mobility problems, and nearly 70% did not drive (Goldstein et al., 2012), both of which serve as additional barriers to travel. These issues are especially difficult to overcome for people who live in rural areas, with increased required travel distance and often nonexistent public transit options (Iezzoni et al., 2006). Recently, the COVID-19 pandemic also had a significant impact on the rate at which patients were able to meet with their eye doctors, with studies observing an increase of over 20% in missed ophthalmology appointments in 2020 (Brant et al., 2021).

Telehealth

A promising direction for addressing the challenges associated with traveling to in-clinic visits is telehealth, the remote connection of patients to their care providers. A study by Portney et al. in 2021 illustrated the potential of telehealth in ophthalmology during the COVID-19 pandemic, finding that some clinics experienced a four-fold increase in telehealth visits when in-person appointments were not possible. Complementing that potential rising demand for telehealth, many new remote tests for both VA and CS have been recently developed. A list of recently developed remote tests is included in Table 4-1, including citations to the studies in which they were described. The majority of the tests listed in Table 4-1 operate as either mobile phone or tablet

Test Name	Test Device	Visual Dimension	Limit of Agreement with Clinical Standard	Citation
AAPOS Vision Screening	Tablet	VA	-0.19 – 0.23	Azis et al., 2019
Eye Chart	Cellphone	VA	-0.21 – 0.19	Tiraset et al., 2021
EyeChartPro	Tablet	VA	-0.30 – 0.30	Zhang et al., 2013
Eye Handbook	Cellphone	VA	NR	Tofigh et al., 2015
Fast Screening VA	Tablet	VA	-0.27 – 0.15	Rodríguez-Vallejo et al., 2016
Home Acuity Test	Mailed Physical Chart	VA	-0.44 – 0.23	Crossland et al., 2021
Handy Eye Check	Tablet	VA	NR	Toner et al., 2014
Odysight	Cellphone or Tablet	VA	-0.3 – 0.24	Brucker et al., 2019
Paxos Checkup	Cellphone	VA	NR	Pathipati et al., 2016
Peek Acuity	Cellphone	VA	-0.44 – 0.55	Bastawrous et al., 2015
Peek Near Vision	Cellphone	VA	-0.22 – 0.24	Katibeh et al., 2022
Sightbook	Tablet	VA	NR	Phung et al., 2016
“Snellen” DrBloggs Ltd.	Cellphone	VA	-.032 – 0.36	Perera et al., 2015
V@home	Cellphone	VA	-0.29 – 0.26	Han et al., 2019
Verana Vision Test	Computer	VA	-0.50 – 0.26	Bellsmith et al., 2022
Visual Acuity XL	Tablet	VA	NR	Black et al., 2013
2020 Duo FLEX VA Chart	Tablet	VA	NR	Manzanaro et al., 2015
Aston CS Test	Tablet	CS	NR	Kingsnorth et al., 2016
Fast Screening CS	Tablet	CS	-0.28 – 0.26	Rodríguez-Vallejo et al., 2016
Manifold AUCSF	Tablet	CS	NR	Crossland et al., 2024
PeekCS	Cellphone	CS	-0.27 – 0.29	Habtamu et al., 2019

PopCSF AUCSF	Tablet	CS	NR	Crossland et al., 2024
Sightbook	Computer	CS	NR	Kim et al., 2022
Spaeth-Richman CS Test	Computer - Web Browser	CS	-0.33 – 0.30	Faria et al., 2015

Table 4-1: Available remote tests of either VA or CS. Limits of Agreement (LoA) are listed in logMAR for VA tests and in logCS for CS tests. LoA are calculated as the mean difference between two tests ± 1.96 times the standard deviations of the differences between the tests, and provide a range within which 95% of differences between the tests can be expected to fall. Additional information about and interpretation of LoAs are included in the Statistical Analysis segment of the Methods section.

applications, with the exceptions of the Home Acuity Test, which is a physical paper test that is mailed to patients (Crossland et al., 2021), and the Spaeth-Richman CS Test, which is accessed via a web browser (Faria et al., 2015).

In addition to their value in ophthalmology, these remote tests could be a valuable tool in remote LV rehabilitation. Studies have demonstrated that telehealth in LV rehabilitation is an effective method for ensuring proper use of magnifiers for reading and staying in contact with patients (Bittner et al., 2018, 2022, 2024), but VA measurement is still a necessary part of the process. Remote testing of VA would alleviate that issue, allowing more rehabilitation services to be administered without the need for travel.

Limitations of Telehealth

Remote testing comes with its own set of challenges, including the abundance of tests lacking validation data. A 2015 study by Perera et al. reviewed 11 remote vision tests available on the Apple app store for mobile devices, and found that there was little standardization of optotype sizes in the app, with errors in size ranging from 7% to 40% (none of those tests are included in Table 4-1 due to their poor size calibration). While

there are tests that have been thoroughly tested and validated, and Table 4-1 can serve as a reference list for them, it can be difficult for someone unfamiliar with current research to select a properly validated test without up-to-date knowledge of remote testing research.

Another potential obstacle for remote testing is patients' discomfort with relying on technology for eye care. A study by Korot et al. in 2022 found that up to 30% of patients are uncomfortable relying on an app for a remote eye exam, and that discomfort was associated with lower remote vision test app usage. That being said, a study by Pidgeon et al. in 2022 found that most patients readily accepted remote eye care delivered by phone or video calls, with no drop in engagement or follow-up visits compared to in-office visits. Notably, telehealth administered by phone or video call involved direct contact with a human care provider, rather than relying on a self-testing using an app, which may have been key in overcoming technology-related discomfort.

Remote VA tests in particular are limited by the range of VA they can evaluate. As of now, the lower bound for remote testing of VA is 1.3 logMAR (Snellen equivalent of 20/400) (Bellsmith et al., 2022; Han et al., 2019; Perera et al., 2015), meaning they are unable to quantify VA for many patients with severe forms of vision loss. This is a significant gap, as patients with severe vision loss are the most likely to experience travel difficulties when commuting to in-person clinic appointments (Swenor et al., 2015), and thus stand to benefit the most from remote testing. Compounding the range issue, most available VA tests present letters in standard contrast polarity (black letters against a white background), but some people with low vision show improved reading speed with

reverse contrast polarity text (white letters against a black background), particularly those with cloudy ocular media (Legge et al., 1985) or retinitis pigmentosa (Ehrlich, 1987). Given this reading performance benefit, reverse contrast polarity may enable VA measurement for those who would otherwise be unable to complete testing due to difficulties recognizing stimuli with standard contrast polarity. However, as of now, reverse contrast polarity options for remote testing are not available.

Prior Findings and Current Aims

In a prior report, which will be referred to as the Chapter 3 Study, we described the development and initial validation of an Online Vision Test (OVT), which is capable of measuring both VA and CS remotely for participants with severe LV (VA poorer than 20/400). In that study, 29 normally sighted participants with artificially restricted vision were tested with both the OVT and standard clinical vision charts in a laboratory setting. We found that the OVT met four goals: 1) it produced results in good agreement with gold standard clinical measures, the ETDRS VA chart and the Pelli-Robson (PR) CS chart; 2) it could be administered online, without the test takers need for any equipment beyond a home computer; 3) it was capable of measuring severely impaired vision, with VA poorer than 20/400; and 4) the tests were easy and quick to administer, and produced results that were straightforward to interpret, with the same output metrics as the corresponding standard charts. We also observed a contrast polarity effect, wherein participants tended to perform better on reverse-polarity versions of both the VA and CS tests.

Expanding upon those findings, the current study continued the validation process with remote testing of participants with naturally occurring low vision. We aimed to answer four primary questions:

1. **Would remote testing either introduce challenges in test administration or reduce the degree of agreement between the OVT and clinical chart-based tests?** Compared to testing in a laboratory setting, at-home remote testing introduces variation in lighting conditions, potential distractions, and potential difficulty in calibration or test setup. As the OVT must be robust to those factors, we aimed to evaluate how they would affect remote administration or agreement with the charts.
2. **Will we observe a consistent contrast polarity effect in a group with more heterogeneous visual conditions?** Reverse contrast can benefit reading for some people with LV, including those with cloudy ocular media, retinitis pigmentosa, and albinism (Legge, 1985; Ehrlich, 1987; Wu et al., 2020). We aimed to determine if the benefit would persist in a group with a more diverse array of diagnoses.
3. **Would the OVT be capable of quantifying the VA for participants with naturally occurring severe vision loss?** Some of our participants had naturally occurring VA poorer than 20/400 – the group for which no existing remote tests can quantify VA. We aimed to determine if we could measure VA in naturally occurring severe LV.
4. **Would there be any signs of discomfort with technology in a generally older low-vision participant sample?** Prior literature has not been conclusive about whether or not more senior populations will be comfortable using modern technology for remote vision testing. Our sample was comprised of older participants, and we aimed to evaluate

whether the technology used in the OVT would lead to any difficulty in the testing process.

While we report on the statistical results of the tests and discuss the degree of agreement between the OVT and chart-based tests, it is important to acknowledge that our current sample size of 14 participants is small. Furthermore, while the OVT was designed with severe LV in mind, several of our participants have mild or moderate LV. These participants were included because the OVT can be useful across the spectrum of LV, from mild to severe, so it is necessary to evaluate the viability of the test across a broad range of vision conditions. To more conclusively validate the test in severe LV, it will be necessary to continue validation with a larger sample, as those with severe LV may have different challenges associated with remote testing.

Methods

Low vision adults underwent two rounds of vision testing, both of which were conducted while they were at their homes. The first round was conducted online and remotely, while the other was conducted with the experimenter in person at the participant's home. The first round was the OVT, which included the o-VA (the online VA test) and the o-CS (the online CS test). All details about the stimuli and mechanical details of the OVT that were used in this round of online testing are identical to those from our prior study, and are described in Chapter 3. The research methods, recruitment materials, and informed consent forms for this study were approved by the Institutional

Review Board of the University of Minnesota, and all research methods adhere to the tenets of the Declaration of Helsinki.

Participants

Participants were low-vision adults recruited via fliers and by direct contact. Those who were contacted directly were individuals who had participated in prior research conducted by our lab. In total, we recruited 14 participants, all of whom completed participation. Participants' mean age was 66, with 3 being male and 11 being female. Participant information, including age, gender, and ocular diagnosis is included in Table 4-2. The only inclusion criteria for participation in the study were having a best corrected VA of poorer than 20/40, being aged 18 or over, and having access to a home computer for remote testing. The mean VA, as measured in this study, was 0.84 logMAR.

Participant #	Age	Gender	Diagnosis	ETDRS VA (logMAR)	Mars CS (logCS)
1	63	F	Familial Exudative Vitreoretinopathy	1.54*	0.56
2	67	M	Glaucoma	1.31*	0.4
3	61	F	Macular Holes	0.84	1.56
4	59	M	Stargardts Disease	1.0	1.16
5	71	F	AMD	0.82	1.48
6	62	F	Diabetic Retinopathy	1.14	1.04
7	62	M	Glaucoma	1.06	1
8	57	F	Pseudo Xanthoma Elasticum	0.54	1.08
9	74	F	Glaucoma	0.8	0.92
10	70	F	Myopic Degeneration	0.42	1.52
11	76	F	Surgical damage	0.42	1.56
12	70	F	Retinal Detachment	0.6	1.6
13	76	F	Retinitis Pigmentosa	0.78	0.92
14	62	F	Glaucoma	0.56	1.28

Table 4-2: Participant demographics, diagnostic information, and vision status. ETDRS VA is reported in logMAR, while Mars CS is reported in logCS. A * next to a VA value indicates the participant had a VA poorer than 20/400.

Participants were compensated at a rate of $15 per hour, and informed consent was obtained from all participants prior to beginning the first round of testing.

Test Setup and Calibration

For the remote testing portion of the study, participants were required to have three tools on hand: their own computer or tablet to take the test on, a tool to measure the size of an image on their screen (e.g. a ruler or tape measure), and an object of known length (which would be used to set their viewing distance for the test). The possible objects of known length were all common household objects: a 12-inch ruler, a paper towel roll/tube, or a standard 8x11 inch sheet of printer paper, were all chosen because they have standardized lengths (regardless of brand), and are common in most homes. The way in which these were used is explained below. Additionally, if participants were unable to read or use their measuring tool, they were asked to have a friend or family member available who would be able to use it for them, which occurred with three participants.

The OVT is carried out in five steps: first, the test administrator runs the testing software, PsychoPy (Peirce et al., 2019), on their own computer. Second, the administrator invites the test taker via an emailed hyperlink to a Zoom conference call. The participant does not need to have Zoom installed, and can instead join via a web browser. Third, once the test taker is in the call, the administrator initiates screen sharing, so the test taker can see the test stimuli that will be presented on the administrator's screen. Fourth, the test is calibrated on the test taker's screen. Fifth, the test taker views

their screen at a specific viewing distance, and verbally responds to on-screen stimuli until the test is completed.

Calibration requires two steps: measuring image size scaling on the participant's screen, and setting viewing distance. As display sizes are typically different between the administrator's screen and the participant's screen, calculation of test results requires accounting for this size difference. Measurement of that difference is done by presenting a screen containing a single white square against a black background. Participants are then asked to measure the width and height of the square on their screen (or to ask a friend or family member to do so). This size information is compared to the size of the square on the administrator's screen to determine how the participant's screen scales the size of images.

Viewing distance calibration is done using the household object of known length, either 11 or 12 inches. The participant is instructed to hold the object to their chin, then slowly move forward until the other end of it touches the center of their monitor. Viewing distance is measured before each test, and participants are instructed to stay at that distance for the duration of each test, with breaks for rest between tests. This method was chosen because prior research has shown that participants were able to maintain a consistent viewing distance during online psychophysical tasks of up to 10 minutes (Li et al., 2020). Additionally, we confirmed in our Chapter 3 Study that participants were able to maintain consistent viewing distance during the OVT tests, which typically take less than 5 minutes to complete.

After completing calibration, participants were instructed on how to complete a Tumbling E procedure by completing practice trials. For the practice trials, an E was presented at the maximum size for the participant's display, and was presented at one of four orientations, chosen at random: 0°, 90°, 180°, or 270°. Participants were told that a capital letter E would appear on their screen, "facing up, down, left, or right," and that they were to verbally state which direction the E was pointing. After the participant responded, the E would disappear, to be replaced with another E at one of the four orientations, chosen at random after 0.5 seconds. Participants were given feedback about whether they were correct or not to be sure they understood the procedure, though feedback was not provided during actual testing. This process continued until they correctly identified an E at all four possible orientations.

VA and CS Testing

VA and CS testing were conducted using the OVT as described in Chapter 3. Both the o-VA and the o-CS are 4AFC Tumbling E style tests, and both utilize a four-down, one-up staircasing progression. All testing was completed binocularly. Before remote testing began, participants were asked to turn off any room lighting and to prevent any direct sunlight or other sources of glare from affecting their screens when possible by closing nearby blinds or curtains and positioning themselves away from windows. Specific illumination conditions of the rooms in which participants completed their online testing were not measured.

In the o-VA, the first E displayed is very large, filling the screen, and with each step down, it decreases in size by 0.1 log unit. The range the o-VA can measure depends on two factors: 1) the screen size it is being displayed on, and 2) viewing distance at which the test is taken. While viewing distance could theoretically have a major impact on the range of testable VAs, in practice, screen size was the more prominent of these two factors because all tests were conducted with viewing distances of either 11 or 12 inches, while screen sizes varied more from one participant to another. For instance, assuming a viewing distance of 11 inches on a 10.2 inch display (a typical size for a small iPad), the test ranges from 0.1 – 2.4 logMAR, while on a 43 inch display (a possible size for a very large computer monitor), the test ranges from 0.78 – 3.08 logMAR.

In the o-CS, the first E displayed has a very high contrast with its background, and with each step down, the contrast between the E and the background decreases such that each step is equivalent to a decrease of 0.1 logCS. The range that the o-CS can evaluate is independent of the screen it is displayed on, and ranges from 0.1 logCS to 2.0 logCS. Contrast variations between each step were created using dither patterns with varying proportions of white and black pixels. By utilizing only white and black pixels, this method avoids nonlinearities in screen luminance that arise from greyscale mapping. Additionally, the angular size of the E stimuli was matched to the angular size of the letters on the Mars chart when viewed from 40 cm, as differences in angular size can affect CS measurement (Njeru et al., 2021).

Chart-Based Testing

For the in-person portion of the study, chart-based testing of VA and CS was conducted at participants' homes using small, portable charts. VA testing was conducted using the New ETDRS chart, while CS testing was conducted using the Mars chart. At-home testing was done wherever there was an overhead light available in the participant's home, typically at a dining room table. While the luminance of the charts was not measured for every testing session, care was taken to avoid casting any shadows or creating any glare on the chart surfaces. As with the OVT, all chart-based testing was conducted binocularly.

The New ETDRS chart, is a miniature, portable version of the standard ETDRS chart. Participants viewed the test from 40 cm, and were asked to start on the top line and read letters as far down on the chart as possible. As they approached their VA thresholds, participants were prompted to continue reading, and to guess if necessary. This continued until they were unable to correctly report any of the letters on a given line, at which point testing stopped. VA was scored by starting at 1.3 logMAR at the top of the chart, and reducing the logMAR value by 0.02 for each letter correctly identified. For the one participant who had VA poorer than 1.3 logMAR, the viewing distance was reduced to 20 cm, and the scoring was adjusted for the change in viewing distance.

CS testing was done with the Mars chart, which is similar in format to the Pelli-Robson chart, but is smaller and portable (Arditi et al., 2005). Participants viewed the chart from 40 cm, and were asked to start on the top line and read letters as far down on

the Mars chart as possible, guessing if necessary. This continued until they missed two letters in a row, at which point testing stopped. CS was scored by starting at 0 logCS and increasing the logCS value by 0.04 logCS for each letter correctly identified.

Statistical Analysis

We evaluated test agreement in two ways. The first was assessing the magnitude of the differences between mean scores on the tests (i.e. bias), which was done using paired samples t-tests. Comparison of mean scores was conducted between the OVT and the charts, as well as between the two contrast polarity versions of both the o-VA and o-CS.

The second way we assessed agreement was by calculating the Limits of Agreement (LoA) between the OVT and the charts (Bland & Altman, 1986), a measure of agreement based on the variance in the differences between the tests. An LoA is an interval within which 95% of the differences between tests are likely to fall, calculated as the mean of the differences ±1.96 standard deviations of the differences. A smaller range between the lower and upper bounds of an LoA corresponds to lower variance, and thus more consistent, predictable differences, but there is no specific range at which the variance is categorized as “acceptable” or “unacceptable.” Instead, the variance of the differences between tests must be interpreted in the context of their use. Therefore, we used the test-retest reliabilities of the ETDRS and PR charts as benchmarks for the variance of gold-standard clinical measures. Test-retest reliabilities of the ETDRS and PR charts was evaluated for groups of LV subjects by Kiser et al. in 2005, who found they

were ±0.2 logMAR for the ETDRS chart and ±0.35 for the PR chart. As that level of variance is acceptable in clinical settings, those values served as a ceiling for the expected variability of differences for our new tests. If the LoA range was smaller than or close to the test-retest reliability range of the charts (i.e. ±.02 logMAR between the o-VA and the ETDRS chart and ±0.35 logCS between the o-CS and the Mars chart), the level of variance was deemed very good, as was the level of agreement. Larger ranges indicated more variance and thus poorer agreement.

Results

Visual Acuity

Mean VA values for each test condition, mean differences between VA values from the o-VA and ETDRS chart, and LoAs between the o-VA and ETDRS chart are listed in Table 4-3. The o-VA was biased toward significantly poorer mean VA scores (i.e. larger logMAR values) than the ETDRS chart in both contrast polarity versions, $t(13) = 1.796$, $p = 0.048$ for standard polarity, and $t(13) = 2.051$, $p = 0.03$ for reverse polarity. In both cases, the differences were equivalent to slightly less than one line on the ETDRS chart. The two contrast polarity versions of the test produced very similar mean scores, with no statistically significant difference between them, $t(13) = 0.366$, $p = 0.360$.

Test	Mean VA (±*SD*)	Difference from ETDRS	LoA Range [Upper Bound, Lower Bound]
ETDRS Chart	0.84 (±0.33)	---	---
Standard Polarity o-VA	0.92 (±0.28)	-0.08*	±0.33 [-0.40, 0.25]
Reverse Polarity o-VA	0.93 (±0.28)	-0.09*	±0.31 [-0.40, 0.22]

Table 4-3: Results from VA Testing. All values are in logMAR. A * indicates the mean VA value of a test was significantly different from the ETDRS chart.

The range between the upper and lower bounds of the LoAs were similar in the standard and reverse polarity versions of the o-VA, indicating the variance in the differences was similar across versions. In both versions, the LoAs covered slightly more than ±0.3 logMAR, corresponding to a factor of two in Snellen notation, or three lines on the ETDRS chart. Figure 4-1 includes two Bland-Altman plots, one for each contrast polarity version, which illustrate both LoAs and the differences between the tests for individual participants. Notably, Figure 4-1 shows that in both polarities, differences between the tests were equivalent to less than 1 line on the ETDRS chart for 6 of 14 participants, and were equivalent to less than 2 lines for 11 of 14 participants. However, the range of the LoAs was greater than the test-retest reliability of the ETDRS chart (±0.2 logMAR), as measured by Kiser et al. (2005), indicating that agreement between the o-VA and the chart was slightly poorer than agreement of the chart with itself.

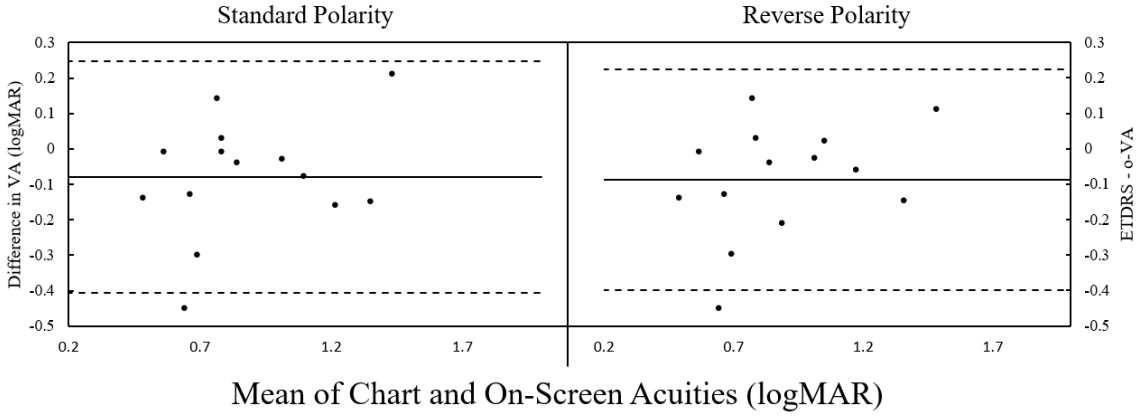


Figure 4-1: A composite of two Bland-Altman plots, one for each o-VA contrast polarity, which illustrate the differences between the o-VA and ETDRS values. The standard polarity version is displayed on the left side, and reverse polarity on the right. The x-axis of each plot is the mean VA of the o-VA and ETDRS tests, while the y-axis is the difference between the tests (ETDRS – o-VA) and each point is an individual participant. LoAs are denoted by the horizontal lines in each plot – the solid line is the mean difference, while the dashed lines are the upper and lower bounds.

Contrast Sensitivity

Mean CS values for each test condition, mean differences between CS values from the o-CS and Mars chart, and LoAs between the o-CS and Mars chart are listed in Table 4-4. The mean score on the standard polarity o-CS was slightly biased toward better CS scores (i.e. higher logCS value) than the mean score on the Mars chart, while the mean reverse polarity score was slightly biased toward worse scores. However, neither of those differences were statistically significant. There was a statistically significant difference between the two contrast polarity versions of the o-CS, with the standard polarity test resulting in a mean score that was 0.18 logCS better than the reverse polarity version, $t(13) = 3.328$, $p = 0.005$.

Test	Mean CS (±*SD*)	Difference from Mars	LoA Range [Upper Bound, Lower Bound]
Mars Chart	1.15 (±0.38)	---	---
Standard Polarity o-CS	1.21 (±0.46)	-0.06	±0.58 [-0.46, 0.7]
Reverse Polarity o-CS	1.03 (±0.48)	0.12	±0.55 [-0.61, 0.49]

Table 4-4: Results from CS Testing. All values are in logCS.

The range between the upper and lower bounds of the LoAs were similar in the standard and reverse polarity versions of the o-CS. LoAs covered approximately ±0.55 logCS, corresponding to slightly more than 14 letters on the Mars chart (slightly less than four steps on the PR chart). Figure 4-2 includes two Bland-Altman plots, one for each contrast polarity version, illustrating both LoAs and differences between tests for individual participants. The range of the LoAs was greater than the test-retest reliability of the chart (±0.35 logCS) (Kiser et al., 2005), indicating that the variance in the differences between the o-CS and the chart was slightly greater than the variance in the differences between the chart and itself. However, it is worth noting that differences were less than 0.35 for 11 of 14 participants in standard polarity, and for 10 of 14 participants in reverse polarity.

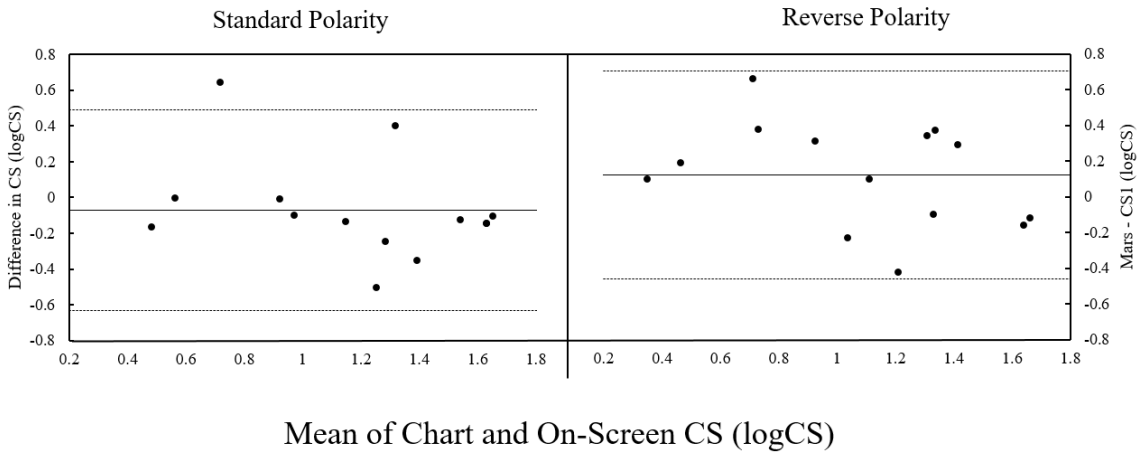


Figure 4-2: A composite of two Bland-Altman plots, one for each o-CS contrast polarity, which illustrate the differences between the o-CS and Mars chart values. The standard polarity version is displayed on the left side, and reverse polarity on the right. The x-axis of each plot is the mean CS of both the o-CS and March tests, while the y-axis is the difference between the tests (Mars – o-CS) and each point is an individual participant. LoAs are denoted by the horizontal lines in each plot – the solid line is the mean difference, while the dashed lines are the upper and lower bounds.

Discussion

In Chapter 3, we assessed agreement between the OVT and clinical chart-based tests in a controlled, laboratory environment. Testing was conducted with normally sighted participants who wore blur goggles that caused either mild or severe restriction of VA and CS. We found generally good agreement between the OVT and the charts, though agreement was worse between VA tests with mild restriction. We also found that performance was generally slightly better on the OVT compared to the charts, and also slightly better in reverse polarity compared to standard polarity. Here, our testing more accurately reflected the real use case conditions for the OVT. Testing was conducted remotely, with participants in the less controlled environment of their homes, and participants had naturally occurring LV with a heterogeneous array of diagnoses. These changes introduce factors such as differences in home lighting, a variety of computer

platforms, potential distractions, and differences in vision status that could affect VA or CS measurement. We aimed to determine how the OVT would perform with those additional challenges, dividing our evaluation into four primary questions.

The first question was whether or not the OVT would continue to both demonstrate good agreement with the charts and be easy to administer, given the differences in testing environment and participant pool. As in the Chapter 3 Study, agreement was evaluated both in terms of the magnitude of the differences between the mean scores of the tests, as well as the level of variance in those differences. The magnitude of the differences is the aspect that reflects bias between the tests, indicating whether performance on one test tends to be better than performance on the other, and if so, how big those differences are. The level of variance in the differences is the aspect that reflects how consistent the differences between the OVT and the charts are. The level of variance is a key part of agreement because if variance is low, it is possible to use a score from the OVT to reliably predict a score on the chart by simply accounting for bias between the tests.

We found that the magnitudes of the differences between the mean scores of the tests were small, which indicated this aspect of agreement was good. The mean scores from both contrast polarities of the o-VA were statistically significantly poorer than those from ETDRS chart, but the differences equated to only approximately 4 letters. This small difference is similar to differences observed in other prominent remote tests of VA such as the Peek Acuity test, which differed from the ETDRS chart by an average of 0.07 logMAR, equating to about 3 letters (Samanta et al., 2023; Bastawrous et al., 2015).

Notably, the poorer mean scores on the o-VA relative to the ETDRS chart was a departure from our findings in the Chapter 3 Study, where scores were better on the o-VA. One potential explanation for this discrepancy is that the home environment may have introduced factors that negatively impacted scores on the o-VA, such as glare on display screens or distractions. While participants were asked to avoid these issues during test setup, and no participants reported either issue arose, we were not able to evaluate the testing environment in their home during o-VA testing. However, while these issues are possible, it is also plausible that they would not affect the majority of test takers in the broader population, and may have had a greater impact due to our relatively small sample size. Therefore, it will be necessary to gather a larger sample to determine if the poorer o-VA scores are due to problems encountered in the homes of a small number of participants, or they were due to systematic differences between the o-VA and ETDRS chart.

Differences were even smaller between the o-CS and Mars chart, equating to approximately 1.5 letters on the Mars chart for standard polarity and 3 letters for reverse polarity, but neither was statistically significant. Overall, the magnitude of the differences was small enough to suggest that while some small degree of bias was present, it can be easily accounted for. Additionally, these mean differences are even smaller than those observed in the Chapter 3 Study, suggesting that the changes in testing environment and participant pool did not introduce any major positive or negative bias in OVT results.

As described in the Statistical Analysis portion of the Methods section, variance in the differences between the tests was evaluated by comparing the range between the

bounds of the LoAs to the test-retest reliability of the charts, which served as benchmarks for variance that is acceptable in a gold-standard clinical measurement tool. We found that the range of the LoAs was greater than the test-retest reliability of the charts. This indicated that with regard to variance in the differences between the tests, the o-VA showed moderate agreement with the ETDRS, while the o-CS showed poor agreement with the Mars chart. Practically speaking, this means that the OVT is less predictive of an individual's chart scores than a prior test using the chart. For instance, if an individual's VA was measured with the ETDRS chart and found to be 20/100, a subsequent measurement with the ETDRS chart would likely be within ±0.2 logMAR, which ranges from approximately 20/60 to 20/160. However, that same individual's VA measured using the o-VA would likely be within ±0.3 logMAR (after accounting for bias), which ranges from 20/50 to 20/200. Thus, while the OVT does effectively provide an estimate of VA and CS, the precision of those estimates is slightly lower than a measurement using the ETDRS or Mars/PR charts. That being said, given the many potential differences home environments and test displays, alongside the access to remote testing that the OVT provides, these data support the utility of the OVT for connecting to patients who would otherwise go without care. This aspect of agreement was poorer than in the Chapter 3 Study, where the variance of the differences were similar to the charts' test-retest reliability in most circumstances.

The variance of the differences was greater in this study compared to the Chapter 3 Study, where it was similar to the charts' test-retest reliability, and this could be due to several factors. There was likely more variance in the lighting conditions across

participants' homes, compared to laboratory testing conditions. This may have affected the light levels on their displays, which could have affected VA and CS results (Buhren et al., 2006; Tidbury et al., 2016). It is also possible that there may have been inaccuracies in calibration. While participants in the Chapter 3 Study were able to effectively measure the calibration square and maintain viewing distance, we were not able to evaluate the measurement accuracy and head stability of participants with low vision in the present study, which could have affected the accuracies of the tests.

Lastly, we note the presence of outliers in our sample with particularly high differences between the OVT and the charts, which drove up the variance of the differences. 2 participants had especially high differences between the o-VA and the ETDRS chart, while 2 different participants had especially high differences between the o-CS and Mars chart. If those outlier participants are removed from the o-VA data, the LoA ranges are reduced to ±0.21 logMAR for standard polarity and ±0.23 logMAR for reverse polarity. Similarly, if the outliers are removed from the o-CS data, the LoA range drops to ±0.33 logCS for standard polarity and ±0.44 logCS for reverse polarity. With those outliers removed, the ranges either meet or approach the threshold of test-retest reliability of the charts. A retrospective review of these outlier participants did not reveal any unusual demographic or diagnostic information about them, making it unclear why exactly their differences were so large. While it cannot be conclusively determined with our current dataset, it is possible that with a larger sample size the effect of these outliers would be reduced, leading to improved consistency.

We also asked whether the less controlled testing environment would introduce unforeseen problems with the testing procedure. We found no surprising issues, with administration of the test proceeding smoothly for all of our testing sessions. We encountered no confusion on the task from participants, and sighted assistance was only necessary for test calibration for three of our participants. Once calibration was completed, each test proceeded quickly, taking two to five minutes each.

Our second question was whether or not we would observe a contrast polarity effect, as was observed in the Chapter 3 Study, where participants typically demonstrated better acuity and contrast sensitivity on tests presented in reverse polarity. In this study, there was not a consistent contrast polarity effect. We found no difference between polarities in the o-VA, but did find that scores were significantly better on the standard polarity o-CS compared to the reverse polarity version (a difference equating to approximately 4 letters on the Mars chart). The better performance with standard polarity is surprising, given our prior finding in Chapter 3 that performance is better with reverse polarity. A possible explanation is that in the Chapter 3 Study, vision was restricted in the same way for all participants: using blur foils that scattered light in a similar fashion to eyes with cloudy ocular media, which benefit from reverse contrast polarity. In this study, our participants had a more heterogeneous array of vision conditions, most of which did not have cloudy ocular media, which may have reduced the consistency of the effect.

Our third question was whether the OVT would be capable of measuring VA poorer than 20/400, as no other currently available remote VA tests is designed to do so. In Chapter 3, we found that we could measure VA artificially restricted to poorer than

20/400, and the results of this study support the same conclusion. One participant had a VA poorer than 20/400, while another had acuity close to 20/400, and their VA was effectively measured by the o-VA. Notably, these participants were not the outliers with especially large differences between the OVT and the charts.

Interestingly, three of our participants reported that their doctors did not numerically quantify their VA, instead classifying their vision as being at the level of "finger counting" or "hand motion." That qualitative description is usually reserved for sight with VA poorer than 20/200, which is the threshold for legal blindness in the United States. As there are no legal vision categories beyond that point, further quantification may have been considered unnecessary. However, we found that participants were generally glad to have been told their Snellen acuity, as opposed to simply being told they were blind. This aligns with findings from prior research, which has found that patients having their vision measured expressed an "overwhelming wish" to receive feedback about their vision test results, with 88% preferring to see their test results immediately after testing (Korot et al., 2022). As the OVT is quick and easy to administer, it could promote better communication of VA levels poorer than 20/400, promoting improved eye care satisfaction for patients with severe vision impairments.

Some prior research has suggested that older participants are not comfortable with technology for remote vision testing. To address this, for our final question we asked: given our older participant group, would there be any signs of discomfort with the technology used in the OVT? While our sample may have been biased due to being comprised of individuals who were willing to sign up for a study about an online vision

test, we found that the technology used for the OVT did not create any major roadblocks in our testing sessions. The majority of our participants having no issues at all, with any problems that did arise being solved within 5 minutes of connecting. Our finding of comfort with technology aligns with findings by Bitner et al. (2024), who found that in a study of telerehabilitation, participants reported being equally comfortable with telerehabilitation as they were with in-office rehabilitation sessions.

Limitations

The most significant limitation of our study is the small sample size, which limits the statistical power of our analysis and inflates the impact of outliers on our results. Another potential limitation is the number of participants we were able to recruit with severely reduced VA. While we did have two that had VA poorer than the 20/400, it would be valuable to demonstrate the efficacy of the o-VA with more participants with VA worse than 20/400. Lastly, while the OVT's use of technology was not a problem for test administration, it may limit the number of people who are able to access it. Some evidence has suggested that nearly half of individuals in the UK older than 75 years old do not regularly use the internet (Crossland et al., 2021). In the US, digital device ownership is another issue for patient access to the OVT, as recent evidence suggests that individuals with LV are significantly less likely to own computers or tablets than their sighted peers (Thomas et al., 2024).

Future Research

There are several directions in which future research could continue to expand upon our work here. Additional testing with more participants with cloudy ocular media could also help to determine if there is an interaction between specific ocular diseases and the benefit of reverse contrast polarity. Additionally, in the Chapter 3 Study, we found that scores were better on the OVT compared to the charts by a small, but consistent margin. This was expected, as the tumbling E test task from the OVT is simpler than the letter recognition task from the charts, and simpler tasks have been found to lead to better vision test scores. However, we did not observe that effect here – in fact, the standard polarity o-CS was the only test condition where mean scores on the OVT were higher than the charts. The magnitude of this difference was fairly small, meaning it is difficult to draw conclusions about the nature of this effect here. However, with a larger sample size, it may be possible to systematically assess how differences among ocular pathologies or other variables might interact with task characteristics to affect test scores.

Conclusion

To conclude, we administered the online vision tests remotely to low-vision participants at their homes, measuring both VA and CS, without requiring any adjustments to interior lighting or any specialized equipment aside from a computer. We found that while agreement with clinical charts was good with regard to mean differences between scores, agreement with regards to variance in the differences between the tests was lower. This meant that while on average test scores are similar, it is difficult to

predict VA or CS scores from clinical charts based on scores from the OVT. However, results reported in Chapter 3 demonstrate that the OVT has the potential to show good agreement with the charts, and a small number of outliers had a large negative impact on agreement. Considering those factors, it is possible that additional data may lead to improvements in agreement. Furthermore, while estimates of chart scores may not be as precise as prior measurements with the same chart, the OVT does provide a reasonable estimate of ETDRS VA and Mars CS. Additionally, the test is valuable as an option that is not limited to any specific operating system, working on any device that can connect to a Zoom call, and is robust to differences across display devices. Given these factors, once additional data is collected, the OVT may prove to be an effective method for remotely assessing both VA and CS, suitable for ophthalmologists to connect to their patients and estimate of their vision status when in-person visits to the clinic are not a viable option.

References

Adeyemo, O., Jeter, P. E., Rozanski, C., Arnold, E., Dalvin, L. A., Swenor, B., ... & PLoVR Study Group. (2017). Living with ultra-low vision: an inventory of self-reported visually guided activities by individuals with profound visual impairment. *Translational Vision Science & Technology*, *6*(3), 10-10. https://doi.org/10.1167/tvst.6.3.10

Alexander, K. R., Derlacki, D. J., & Fishman, G. A. (1995). Visual acuity vs letter contrast sensitivity in retinitis pigmentosa. *Vision research*, *35*(10), 1495-1499. https://doi.org/10.1016/0042-6989(95)98729-S

Arditi, A. (2005). Improving the design of the letter contrast sensitivity test. *Investigative ophthalmology & visual science*, *46*(6), 2225-2229. https://doi.org/10.1167/iovs.04-1198

Aslam, T. M., Parry, N. R., Murray, I. J., Salleh, M., Col, C. D., Mirza, N., ... & Tahir, H. J. (2016). 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 Archive for Clinical and Experimental Ophthalmology*, *254*, 891-899. https://doi.org/10.1007/s00417-016-3293-2

Aslam, M. (2023). Cochran's Q test for analyzing categorical data under uncertainty. *Journal of Big Data*, *10*(1), 147. https://doi.org/10.1186/s40537-023-00823-3

Azis, N. N. N., Chew, F. L. M., Rosland, S. F., Ramlee, A., & Che-Hamzah, J. (2019). Parents' performance using the AAPOS Vision Screening App to test visual acuity in Malaysian preschoolers. *Journal of American Association for Pediatric Ophthalmology and Strabismus*, *23*(5), 268-e1. https://doi.org/10.1016/j.jaapos.2019.01.019

Bach, M. (1996). The Freiburg Visual Acuity test—automatic measurement of visual acuity. *Optometry and vision science*, *73*(1), 49-53.

Bach, M. (2006). The Freiburg Visual Acuity Test-variability unchanged by post-hoc re-analysis. *Graefe's Archive for Clinical and Experimental Ophthalmology*, *245*, 965-971. https://doi.org/10.1007/s00417-006-0474-4

Bailey, I. L., Jackson, A. J., Minto, H., Greer, R. B., & Chu, M. A. (2012). The Berkeley rudimentary vision test. *Optometry and Vision Science*, *89*(9), 1257-1264. doi: 10.1097/OPX.0b013e318264e85a

Bailey, I. L., & Lovie-Kitchin, J. E. (2013). Visual acuity testing. From the laboratory to the clinic. *Vision research*, *90*, 2-9. https://doi.org/10.1016/j.visres.2013.05.004

Barhorst-Cates, E. M., Rand, K. M., & Creem-Regehr, S. H. (2019). Navigating with peripheral field loss in a museum: Learning impairments due to environmental

complexity. *Cognitive Research: Principles and Implications*, *4*, 1-10. https://doi.org/10.1186/s41235-019-0189-9

Bastawrous, A., Rono, H. K., Livingstone, I. A., Weiss, H. A., Jordan, S., Kuper, H., & Burton, M. J. (2015). Development and validation of a smartphone-based visual acuity test (peek acuity) for clinical practice and community-based fieldwork. *JAMA ophthalmology*, *133*(8), 930-937. doi: 10.1001/jamaophthalmol.2015.1468

Bellmann, C., Unnebrink, K., Rubin, G. S., Miller, D., & Holz, F. G. (2003). Visual acuity and contrast sensitivity in patients with neovascular age-related macular degeneration: results from the Radiation Therapy for Age-Related Macular Degeneration (RAD-) Study. *Graefe's archive for clinical and experimental ophthalmology*, *241*, 968-974. https://doi.org/10.1007/s00417-003-0689-6

Bellsmith, K. N., Gale, M. J., Yang, S., Nguyen, I. B., Prentiss, C. J., Nguyen, L. T., ... & Thomas, M. (2022). Validation of home visual acuity tests for telehealth in the COVID-19 era. *JAMA ophthalmology*, *140*(5), 465-471. doi: 10.1001/jamaophthalmol.2022.0396

Bibby, S. A., Maslin, E. R., McIlraith, R., & Soong, G. P. (2007). Vision and self-reported mobility performance in patients with low vision. *Clinical and Experimental Optometry*, *90*(2), 115-123. https://doi.org/10.1111/j.1444-0938.2007.00120.x

Bittner, A. K., Yoshinaga, P., Bowers, A., Shepherd, J. D., Succar, T., & Ross, N. C. (2018). Feasibility of telerehabilitation for low vision: satisfaction ratings by providers and patients. *Optometry and Vision Science*, *95*(9), 865-872. doi*:* 10.1097/OPX.0000000000001260

Bittner, A. K., Kaminski, J. E., Ross, N. C., Shepherd, J. D., Thoene, S. J., Bui, S. Z., ... & BeST-AID Study Team. (2022). Telerehabilitation training to facilitate improved reading ability with new magnification devices for low vision. *Optometry and Vision Science*, *99*(10), 743-749. doi*:* 10.1097/OPX.0000000000001944

Bittner, A. K., Kaminski, J. E., Yoshinaga, P. D., Shepherd, J. D., Chan, T. L., Malkin, A. G., ... & BeST-AID Study Team. (2024). Outcomes of Telerehabilitation Versus In-Office Training With Magnification Devices for Low Vision: A Randomized Controlled Trial. *Translational Vision Science & Technology*, *13*(1), 6-6. https://doi.org/10.1167/tvst.13.1.6

Black, A. A., Wood, J. M., & Lovie-Kitchin, J. E. (2011). Inferior field loss increases rate of falls in older adults with glaucoma. *Optometry and Vision Science*, *88*(11), 1275-1282. doi: 10.1097/OPX.0b013e31822f4d6a

Black, J. M., Jacobs, R. J., Phillips, G., Chen, L., Tan, E., Tran, A., & Thompson, B. (2013). An assessment of the iPad as a testing platform for distance visual acuity

in adults. *BMJ open*, *3*(6), e002730. https://doi.org/10.1136/bmjopen-2013-002730

Bland, J. M., & Altman, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet*, *327*(8476), 307-310. https://doi.org/10.1016/S0140-6736(86)90837-8

Bochsler, T. M., Legge, G. E., Gage, R., & Kallie, C. S. (2013). Recognition of ramps and steps by people with low vision. *Investigative Ophthalmology & Visual Science*, *54*(1), 288-294. https://doi.org/10.1167/iovs.12-10461

Boon, M. Y., Chu, B. S., Lee, P. C., Chiang, T. J., Alshamli, N., Alghamdi, W., ... & Bridge, C. (2015). Perceptions of older people regarding their vision and incident causation. *Optometry and vision science*, *92*(10), 995-1002. doi: 10.1097/OPX.0000000000000695

Bourne, R. R., Flaxman, S. R., Braithwaite, T., Cicinelli, M. V., Das, A., Jonas, J. B., ... & Zheng, Y. (2017). Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. *The Lancet Global Health*, *5*(9), e888-e897. https://doi.org/10.1016/S2214-109X(17)30293-0

Brant, A. R., Pershing, S., Hess, O., Rooney, D., Goldberg, J., Tabin, G., & Wang, S. Y. (2021). The impact of COVID-19 on missed ophthalmology clinic visits. *Clinical Ophthalmology*, 4645-4657. https://doi.org/10.2147/OPTH.S341739

Brucker, J., Bhatia, V., Sahel, J. A., Girmens, J. F., & Mohand-Saïd, S. (2019). Odysight: a mobile medical application designed for remote monitoring—a prospective study comparison with standard clinical eye tests. *Ophthalmology and therapy*, *8*, 461-476. https://doi.org/10.1007/s40123-019-0203-9

Brussee, T., van Nispen, R. M., & van Rens, G. H. (2017). Visual and personal characteristics are associated with reading performance in normally sighted adults. *Clinical and Experimental Optometry*, *100*(3), 270-277. https://doi.org/10.1111/cxo.12482

Bühren, J., Terzi, E., Bach, M., Wesemann, W., & Kohnen, T. (2006). Measuring contrast sensitivity under different lighting conditions: comparison of three tests. *Optometry and Vision Science*, *83*(5), 290-298. doi*:* 10.1097/01.opx.0000216100.93302.2d

Chaikitmongkol, V., Nanegrungsunk, O., Patikulsila, D., Ruamviboonsuk, P., & Bressler, N. M. (2018). Repeatability and agreement of visual acuity using the ETDRS number chart, Landolt C chart, or ETDRS alphabet chart in eyes with or without sight-threatening diseases. *JAMA ophthalmology*, *136*(3), 286-290. doi: 10.1001/jamaophthalmol.2017.6290

Chan, T., Friedman, D. S., Bradley, C., & Massof, R. (2018). Estimates of incidence and prevalence of visual impairment, low vision, and blindness in the United States. *JAMA ophthalmology*, *136*(1), 12-19. doi: 10.1001/jamaophthalmol.2017.4655

Chung, S. T., & Legge, G. E. (2016). Comparing the shape of contrast sensitivity functions for normal and low vision. *Investigative ophthalmology & visual science*, *57*(1), 198-207. https://doi.org/10.1167/iovs.15-18084

Chung, D. C., McCague, S., Yu, Z. F., Thill, S., DiStefano-Pappas, J., Bennett, J., ... & High, K. A. (2018). Novel mobility test to assess functional vision in patients with inherited retinal dystrophies. *Clinical & experimental ophthalmology*, *46*(3), 247-259. https://doi.org/10.1111/ceo.13022

Crossland, M. D., Dekker, T. M., Hancox, J., Lisi, M., Wemyss, T. A., & Thomas, P. B. (2021). Evaluation of a home-printable vision screening test for telemedicine. *JAMA ophthalmology*, *139*(3), 271-277. doi: 10.1001/jamaophthalmol.2020.5972

Crossland, M. D., Dekker, T. M., Dahlmann-Noor, A., & Jones, P. R. (2024). Can children measure their own vision? A comparison of three new contrast sensitivity tests. *Ophthalmic and Physiological Optics*, *44*(1), 5-16. https://doi.org/10.1111/opo.13230

Dev, M. K., Wood, J. M., & Black, A. A. (2021). The effect of low light levels on postural stability in older adults with age-related macular degeneration. *Ophthalmic and Physiological Optics*, *41*(4), 853-863. https://doi.org/10.1111/opo.12827

DiMattina, C. (2022). Luminance texture boundaries and luminance step boundaries are segmented using different mechanisms. *Vision research*, *190*, 107968. https://doi.org/10.1016/j.visres.2021.107968

Donoghue, O. A., Ryan, H., Duggan, E., Finucane, C., Savva, G. M., Cronin, H., ... & Kenny, R. A. (2014). Relationship between fear of falling and mobility varies with visual function among older adults. *Geriatrics & gerontology international*, *14*(4), 827-836. https://doi.org/10.1111/ggi.12174

Dorr, M., Lesmes, L. A., Lu, Z. L., & Bex, P. J. (2013). Rapid and reliable assessment of the contrast sensitivity function on an iPad. *Investigative ophthalmology & visual science*, *54*(12), 7266-7273. https://doi.org/10.1167/iovs.13-11743

Ehrlich, D. (1987). A comparative study in the use of closed-circuit television reading machines and optical aids by patients with retinitis pigmentosa and maculopathy. *Ophthalmic and Physiological Optics*, *7*(3), 293-302. https://doi.org/10.1111/j.1475-1313.1987.tb00746.x

Elliott, D. B., Bullimore, M. A., & Bailey, I. L. (1991). Improving the reliability of the Pelli-Robson contrast sensitivity test. *Clinical vision sciences*, *6*(6), 471-475. https://doi.org/10.1364/NAVS.1991.MD12

Elliot, D. B., Patla, A. E., Flanagan, J. G., Spaulding, S., Rietdyk, S., Strong, G., & Brown, S. (1995). The Waterloo Vision and Mobility Study: postural control strategies in subjects with ARM. *Ophthalmic and physiological Optics*, *15*(6), 553-559. https://doi.org/10.1046/j.1475-1313.1995.95000259.x

Elliott, D. B., & Situ, P. (1998). Visual acuity versus letter contrast sensitivity in early cataract. *Vision Research*, *38*(13), 2047-2052. https://doi.org/10.1016/S0042-6989(97)00382-9

Enoch, J., Jones, L., Taylor, D. J., Bronze, C., Kirwan, J. F., Jones, P. R., & Crabb, D. P. (2020). How do different lighting conditions affect the vision and quality of life of people with glaucoma? A systematic review. *Eye*, *34*(1), 138-154. https://doi.org/10.1038/s41433-019-0679-5

Eshraghi, H., Sanvicente, C. T., Gogte, P., Waisbourd, M., Lee, D., Manzi, R. R., ... & Spaeth, G. L. (2019). Measuring contrast sensitivity in specific areas of vision–a meaningful way to assess quality of life and ability to perform daily activities in glaucoma. *Ophthalmic Epidemiology*, *26*(5), 301-310. https://doi.org/10.1080/09286586.2019.1616773

Faria, B. M., Duman, F., Zheng, C. X., Waisbourd, M., Gupta, L., Ali, M., ... & Spaeth, G. L. (2015). Evaluating contrast sensitivity in age-related macular degeneration using a novel computer-based test, the spaeth/richman contrast sensitivity test. *Retina*, *35*(7), 1465-1473. doi: 10.1097/IAE.0000000000000474

Figueiro, M. G., Plitnick, B., Rea, M. S., Gras, L. Z., & Rea, M. S. (2011). Lighting and perceptual cues: effects on gait measures of older adults at high and low risk for falls. *BMC geriatrics*, *11*, 1-10. https://doi.org/10.1186/1471-2318-11-49

Freedman, A., Achtemeier, J., Baek, Y., & Legge, G. E. (2019). Gaze behavior during navigation with reduced acuity. *Experimental eye research*, *183*, 20-28. https://doi.org/10.1016/j.exer.2018.11.002

Fuhr, P. S., Liu, L., & Kuyk, T. K. (2007). Relationships between feature search and mobility performance in persons with severe visual impairment. *Optometry and Vision Science*, *84*(5), 393-400. doi: 10.1097/OPX.0b013e31804f5afb

Geruschat, D. R., Turano, K. A., & Stahl, J. W. (1998). Traditional measures of mobility performance and retinitis pigmentosa. *Optometry and Vision Science*, *75*(7), 525-537.

Geruschat, D. R., Bittner, A. K., & Dagnelie, G. (2012). Orientation and mobility assessment in retinal prosthetic clinical trials. *Optometry and Vision Science*, *89*(9), 1308-1315.doi: 10.1097/OPX.0b013e3182686251

Ghodrati, M., Morris, A. P., & Price, N. S. C. (2015). The (un) suitability of modern liquid crystal displays (LCDs) for vision research. *Frontiers in Psychology*, *6*, 135053. https://doi.org/10.3389/fpsyg.2015.00303

Goldstein, J. E., Massof, R. W., Deremeik, J. T., Braudway, S., Jackson, M. L., Kehler, K. B., ... & Low Vision Research Network Study Group. (2012). Baseline traits of low vision patients served by private outpatient clinical centers in the United States. *Archives of Ophthalmology*, *130*(8), 1028-1037. doi: 10.1001/archophthalmol.2012.1197

Grech, M., Stuart, T., Williams, L., Chen, C., & Loetscher, T. (2017). The Mobility assessment course for the Diagnosis of spatial neglect: Taking a step Forward?. *Frontiers in neurology*, *8*, 563. https://doi.org/10.3389/fneur.2017.00563

Guerry, E., Caumon, C., Bécheras, E., & Zissis, G. (2021). Influence of chromatic and lighting on the visual environment of the elderly: A critical literature review. *Color Research & Application*, *46*(1), 117-124. https://doi.org/10.1002/col.22562

Habtamu, E., Bastawrous, A., Bolster, N. M., Tadesse, Z., Callahan, E. K., Gashaw, B., ... & Burton, M. J. (2019). Development and validation of a smartphone-based contrast sensitivity test. *Translational vision science & technology*, *8*(5), 13-13. https://doi.org/10.1167/tvst.8.5.13

Han, X., Scheetz, J., Keel, S., Liao, C., Liu, C., Jiang, Y., ... & He, M. (2019). Development and validation of a smartphone-based visual acuity test (vision at home). *Translational Vision Science & Technology*, *8*(4), 27-27. https://doi.org/10.1167/tvst.8.4.27

Hassan, S. E., Hicks, J. C., Lei, H., & Turano, K. A. (2007). What is the minimum field of view required for efficient navigation?. *Vision research*, *47*(16), 2115-2123. https://doi.org/10.1016/j.visres.2007.03.012

Hegde, A. L. (2017). Environmental Lighting in Nursing Homes: A Comparison of Agency Standards that Regulate Nursing Homes with Industry ANSI/IES RP-28 Lighting Standards. *The International Journal of Design in Society*, *12*(1), 1. doi: 10.18848/2325-1328/CGP/v12i01/1-16

Hegde, P. G., Rao, D., Prudhvi, B., & Hegde, N. (2023). Digital contrast sensitivity chart with varying visual acuity: Development and validation. *Oman Journal of Ophthalmology*, *16*(3), 467-471. doi: 10.4103/ojo.ojo_260_22

Ho, A. C., Kleinman, D. M., Lum, F. C., Heier, J. S., Lindstrom, R. L., Orr, S. C., ... & Pollack, J. S. (2020). Baseline visual acuity at wet AMD diagnosis predicts long-term vision outcomes: an analysis of the IRIS Registry. *Ophthalmic Surgery, Lasers and Imaging Retina*, *51*(11), 633-639. https://doi.org/10.3928/23258160-20201104-05

Iezzoni, L. I., Killeen, M. B., & O'Day, B. L. (2006). Rural residents with disabilities confront substantial barriers to obtaining primary care. *Health services research*, *41*(4p1), 1258-1275. https://doi.org/10.1111/j.1475-6773.2006.00534.x

Illuminating Engineering Society. (2016). *Lighting and the visual environment for senior living.* Illuminating Engineering Society of North America.

Jenness, J. W., & Singer, J. (2008). Visibility and conspicuity of detectable warnings for pedestrians with visual impairments. *Transportation Research Record*, *2073*(1), 104-113. https://doi.org/10.3141/2073-12

Jeter, P. E., Rozanski, C., Massof, R., Adeyemo, O., Dagnelie, G., & PLOVR Study Group. (2017). Development of the ultra-low vision visual functioning questionnaire (ULV-VFQ). *Translational Vision Science & Technology*, *6*(3), 11-11. https://doi.org/10.1167/tvst.6.3.11

Kallie, C. S., Legge, G. E., & Yu, D. (2012). Identification and detection of simple 3D objects with severely blurred vision. *Investigative Ophthalmology & Visual Science*, *53*(13), 7997-8005. https://doi.org/10.1167/iovs.12-10013

Katibeh, M., Sanyam, S. D., Watts, E., Bolster, N. M., Yadav, R., Roshan, A., ... & Bastawrous, A. (2022). Development and validation of a digital (Peek) near visual acuity test for clinical practice, community-based survey, and research. *Translational vision science & technology*, *11*(12), 18-18. https://doi.org/10.1167/tvst.11.12.18

Kersten, D., Shakespeare, R., & Thompson, W. (2013). Predicting Visibility in Designs of Public Spaces.

Kim, D. G., Webel, A. D., Blumenkranz, M. S., Kim, Y., Yang, J. H., Yu, S. Y., ... & Myung, D. (2022). A Smartphone-Based Near-Vision Testing System: Design, Accuracy, and Reproducibility Compared With Standard Clinical Measures. *Ophthalmic Surgery, Lasers and Imaging Retina*, *53*(2), 79-84. https://doi.org/10.3928/23258160-20220121-05

Kingdom, F. A., Prins, N. (2016). Adaptive methods. *Psychophysics (Second edition)* (pp. 119 – 148). Academic Press. https://doi.org/10.1016/B978-0-12-407156-8.00005-0

Kingsnorth, A., Drew, T., Grewal, B., & Wolffsohn, J. S. (2016). Mobile app Aston contrast sensitivity test. *Clinical and Experimental Optometry*, *99*(4), 350-355. https://doi.org/10.1111/cxo.12362

Kiser, A. K., Mladenovich, D., Eshraghi, F., Bourdeau, D., & Dagnelie, G. (2005). Reliability and consistency of visual acuity and contrast sensitivity measures in advanced eye disease. *Optometry and vision science*, *82*(11), 946-954. doi*:* 10.1097/01.opx.0000187863.12609.7b

Kitchin, J. E., & Bailey, I. (1981). Task complexity and visual acuity in senile macular degeneration. *The Australian Journal of Optometry*, *64*(6), 235-242. https://doi.org/10.1111/j.1444-0938.1981.tb03013.x

Klein, B. E., Klein, R., Lee, K. E., & Cruickshanks, K. J. (1998). Performance-based and self-assessed measures of visual function as related to history of falls, hip fractures, and measured gait time: the Beaver Dam Eye Study. *Ophthalmology*, *105*(1), 160-164. https://doi.org/10.1016/S0161-6420(98)91911-X

Kollbaum, P. S., Jansen, M. E., Kollbaum, E. J., & Bullimore, M. A. (2014). Validation of an iPad test of letter contrast sensitivity. *Optometry and Vision Science*, *91*(3), 291-296. doi: 10.1097/OPX.0000000000000158

Korot, E., Pontikos, N., Drawnel, F. M., Jaber, A., Fu, D. J., Zhang, G., ... & Balaskas, K. (2022). Enablers and barriers to deployment of smartphone-based home vision monitoring in clinical practice settings. *JAMA ophthalmology*, *140*(2), 153-160. doi: 10.1001/jamaophthalmol.2021.5269

Kumaran, N., Ali, R. R., Tyler, N. A., Bainbridge, J. W., Michaelides, M., & Rubin, G. S. (2020). Validation of a vision-guided mobility assessment for RPE65-associated retinal dystrophy. *Translational Vision Science & Technology*, *9*(10), 5-5. https://doi.org/10.1167/tvst.9.10.5

Kuyk, T., Elliott, J. L., Biehl, J., & Fuhr, P. S. (1996). Environmental variables and mobility performance in adults with low vision. *Journal of the American Optometric Association*, *67*(7), 403-409.

Kuyk, T., Elliott, J. L., & Fuhr, P. S. (1998). Visual correlates of obstacle avoidance in adults with low vision. *Optometry and Vision Science*, *75*(3), 174-182.

Kuyk, T., Elliott, J. L., & Fuhr, P. S. (1998). Visual correlates of mobility in real world settings in older adults with low vision. *Optometry and Vision Science*, *75*(7), 538-547. doi: 10.1097/00006324-199807000-00023

Kuyk, T., & Elliott, J. L. (1999). Visual factors and mobility in persons with age-related macular degeneration. *Journal of Rehabilitation Research and Development*, *36*(4).

Legge, G. E., Rubin, G. S., Pelli, D. G., & Schleske, M. M. (1985). Psychophysics of reading—II. Low vision. *Vision research*, *25*(2), 253-265. https://doi.org/10.1016/0042-6989(85)90118-X

Legge, G. E., Yu, D., Kallie, C. S., Bochsler, T. M., & Gage, R. (2010). Visual accessibility of ramps and steps. *Journal of vision*, *10*(11), 8-8. https://doi.org/10.1167/10.11.8

Legge, G. E. (2016). Reading digital with low vision. *Visible language*, *50*(2), 102.

Li, Q., Joo, S. J., Yeatman, J. D., & Reinecke, K. (2020). Controlling for participants' viewing distance in large-scale, psychophysical online experiments using a virtual chinrest. *Scientific reports*, *10*(1), 904. https://doi.org/10.1038/s41598-019-57204-1

Lukman, A. L., Bridge, C., Dain, S. J., & Boon, M. Y. (2020). Luminance contrast of accessible tactile indicators for people with visual impairment. *Ergonomics in design*, *28*(2), 4-15. https://doi.org/10.1177/1064804619841841

Manandhar, S., Lukman, A., Dain, S. J., Bridge, C. E., Relf, M., & Boon, M. Y. (2022). Luminance contrast preferences of people with a vision impairment for elements in the built environment. *Work*, *73*(4), 1265-1278. doi: 10.3233/WOR-210997

Manzano, A. A., & Lagamayo, M. A. N. (2015). A comparison of distance visual acuity testing using a standard ETDRS chart and a tablet device. *Philipp J Ophthalmol*, *40*, 88-92.

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, *12*(2), 153-157.

Mena-Guevara, K. J., Piñero, D. P., & de Fez, D. (2021). Validation of digital applications for evaluation of visual parameters: a narrative review. *Vision*, *5*(4), 58. https://doi.org/10.3390/vision5040058

National Institute of Building Sciences. (2013). *Design Guidelines for the Visual Environment.* https://www.nibs.org/files/pdfs/NIBS_LVDP_Guidelines_2015.pdf

Njeru, S. M., Osman, M., & Brown, A. M. (2021). The effect of test distance on visual contrast sensitivity measured using the Pelli-Robson chart. *Translational Vision Science & Technology*, *10*(2), 32-32. https://doi.org/10.1167/tvst.10.2.32

Odell, N. V., Leske, D. A., Hatt, S. R., Adams, W. E., & Holmes, J. M. (2008). The effect of Bangerter filters on optotype acuity, Vernier acuity, and contrast sensitivity. *Journal of American Association for Pediatric Ophthalmology and Strabismus*, *12*(6), 555-559. https://doi.org/10.1016/j.jaapos.2008.04.012

Onoja, O. P., Eze, U. A., Kator, C., & Lawan, A. (2023). Association between contrast sensitivity and disease severity in primary open-angle glaucoma. *Nigerian Journal of Ophthalmology*, *31*(3), 77-82. doi: 10.4103/njo.njo_6_23

Pang, Y., Sparschu, L., & Wang, J. (2021). Repeatability of an automated ETDRS contrast threshold measurement. *Ophthalmic and Physiological Optics*, *41*(4), 896-899. https://doi.org/10.1111/opo.12829

Pathipati, A. S., Wood, E. H., Lam, C. K., Sáles, C. S., & Moshfeghi, D. M. (2016). Visual acuity measured with a smartphone app is more accurate than Snellen testing by emergency department providers. *Graefe's archive for clinical and experimental ophthalmology*, *254*, 1175-1180. https://doi.org/10.1007/s00417-016-3291-4

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., ... & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior research methods*, *51*, 195-203. https://doi.org/10.3758/s13428-018-01193-y

Perera, C., Chakrabarti, R., Islam, F. M. A., & Crowston, J. (2015). The Eye Phone Study: reliability and accuracy of assessing Snellen visual acuity using smartphone technology. *Eye*, *29*(7), 888-894. https://doi.org/10.1038/eye.2015.60

Pérez Roche, M. T., Yam, J. C., Liu, H., Gutierrez, D., Pham, C., Balasanyan, V., ... & TrackAI Research Consortium. (2022). Visual acuity and contrast sensitivity in preterm and full-term children using a novel digital test. *Children*, *10*(1), 87. https://doi.org/10.3390/children10010087

Phung, L., Gregori, N. Z., Ortiz, A., Shi, W., & Schiffman, J. C. (2016). Reproducibility and comparison of visual acuity obtained with sightbook mobile application to near card and Snellen chart. *Retina*, *36*(5), 1009-1020. doi*:* 10.1097/IAE.0000000000000818

Pidgeon, J. H., Bhardwaj, M. K., Titterington, P., Latulippe, K., Roh, S., & Ramsey, D. J. (2022). Assessing optometric care delivered by telehealth during the COVID-19 public health emergency. *Therapeutic Advances in Ophthalmology*, *14*, 25158414221123526. https://doi.org/10.1177/25158414221123526

Portney, D. S., Zhu, Z., Chen, E. M., Steppe, E., Chilakamarri, P., Woodward, M. A., ... & Parikh, R. (2021). COVID-19 and use of teleophthalmology (CUT Group): trends and diagnoses. *Ophthalmology*, *128*(10), 1483-1485. https://doi.org/10.1016/j.ophtha.2021.02.010

Richman, J., Zangalli, C., Lu, L., Wizov, S. S., Spaeth, E., & Spaeth, G. L. (2015). The Spaeth/Richman contrast sensitivity test (SPARCS): design, reproducibility and ability to identify patients with glaucoma. *British Journal of Ophthalmology*, *99*(1), 16-20. https://doi.org/10.1136/bjophthalmol-2014-305223

Rodríguez-Vallejo, M., Remón, L., Monsoriu, J. A., & Furlan, W. D. (2015). Designing a new test for contrast sensitivity function measurement with iPad. *Journal of Optometry*, *8*(2), 101-108. https://doi.org/10.1016/j.optom.2014.06.003

Rodríguez-Vallejo, M., Llorens-Quintana, C., Furlan, W. D., & Monsoriu, J. A. (2016). Visual acuity and contrast sensitivity screening with a new iPad application. *Displays*, *44*, 15-20. https://doi.org/10.1016/j.displa.2016.06.001

Samanta, A., Mauntana, S., Barsi, Z., Yarlagadda, B., & Nelson, P. C. (2023). Is your vision blurry? A systematic review of home-based visual acuity for telemedicine. *Journal of telemedicine and telecare*, *29*(2), 81-90. https://doi.org/10.1177/1357633X20970398

Seeger, P., & Gabrielsson, A. (1968). Applicability of the Cochran Q test and the F test for statistical analysis of dichotomous data for dependant samples. *Psychological Bulletin*, *69*(4), 269. https://doi.org/10.1037/h0025667

Sheskin, D. J. (2011). *Handbook of parametric and nonparametric statistical procedures* (5th ed.). Boca Raton, FL: Chapman & Hall/CRC Press.

Sinoo, M. M., Van Tilborg, M., Schols, J. M., & Kort, H. S. (2014). Age related visual pathologies among nursing home residents: an evaluation of light conditions and recording in client files. *World Journal of Social Science Research*, *1*(2), 27-40.

Smith, A. R. (1995). Gamma correction. *Computer Graphics (CG) Memos, Technical Memo*, *9*.

Swenor, B. K., Simonsick, E. M., Ferrucci, L., Newman, A. B., Rubin, S., Wilson, V., & Health, Aging and Body Composition Study. (2015). Visual impairment and incident mobility limitations: the health, aging and body composition study. *Journal of the American Geriatrics Society*, *63*(1), 46-54. https://doi.org/10.1111/jgs.13183

Thomas, J., Almidani, L., Swenor, B. K., & Varadaraj, V. (2024). Digital Technology Use Among Older Adults With Vision Impairment. *JAMA ophthalmology*. doi: 10.1001/jamaophthalmol.2024.0467

Thompson, W. B., Shakespeare, R. A., Liu, S., Creem-Regehr, S. H., Kersten, D. J., & Legge, G. E. (2022). Evaluating the Visibility of Architectural Features for People with Low Vision–A Quantitative Approach. *Leukos*, *18*(2), 154-172. https://doi.org/10.1080/15502724.2021.1890115

Tidbury, L. P., Czanner, G., & Newsham, D. (2016). Fiat Lux: the effect of illuminance on acuity testing. *Graefe's Archive for Clinical and Experimental Ophthalmology*, *254*, 1091-1097. https://doi.org/10.1007/s00417-016-3329-7

Tiraset, N., Poonyathalang, A., Padungkiatsagul, T., Deeyai, M., Vichitkunakorn, P., & Vanikieti, K. (2021). Comparison of visual acuity measurement using three methods: standard ETDRS chart, near chart and a smartphone-based eye chart application. *Clinical Ophthalmology*, 859-869. http://doi.org/10.2147/OPTH.S304272

Tofigh, S., Shortridge, E., Elkeeb, A., & Godley, B. F. (2015). Effectiveness of a smartphone application for testing near visual acuity. *Eye*, *29*(11), 1464-1468. https://doi.org/10.1038/eye.2015.138

Toner, K. N., Lynn, M. J., Candy, T. R., & Hutchinson, A. K. (2014). The Handy Eye Check: a mobile medical application to test visual acuity in children. *Journal of American Association for Pediatric Ophthalmology and Strabismus*, *18*(3), 258-260. https://doi.org/10.1016/j.jaapos.2014.01.011

Treacy, M. P., Hurst, T. P., Conway, M., Duignan, E. S., Dimitrov, B. D., Brennan, N., & Cassidy, L. (2015). The early treatment in diabetic retinopathy study chart compared with the tumbling-E and Landolt-C. *Ophthalmology*, *122*(5), 1062-1063. https://doi.org/10.1016/j.ophtha.2014.11.024

Varadaraj, V., Assi, L., Gajwani, P., Wahl, M., David, J., Swenor, B. K., & Ehrlich, J. R. (2021). Evaluation of tablet-based tests of visual acuity and contrast sensitivity in older adults. *Ophthalmic epidemiology*, *28*(4), 293-300. https://doi.org/10.1080/09286586.2020.1846758

Wai, K. M., Vingopoulos, F., Garg, I., Kasetty, M., Silverman, R. F., Katz, R., ... & Miller, J. B. (2022). Contrast sensitivity function in patients with macular disease and good visual acuity. *British Journal of Ophthalmology*, *106*(6), 839-844. https://doi.org/10.1136/bjophthalmol-2020-318494

Wertheim, A. H. (2010). Visual conspicuity: a new simple standard, its reliability, validity and applicability. *Ergonomics*, *53*(3), 421-442. https://doi.org/10.1080/00140130903483705

West, S. K., Rubin, G. S., Broman, A. T., Munoz, B., Bandeen-Roche, K., Turano, K., & SEE Project Team. (2002). How does visual impairment affect performance on tasks of everyday life?: The SEE Project. *Archives of Ophthalmology*, *120*(6), 774-780. doi: 10.1001/archopht.120.6.774

Woodward, M. A., Jeganathan, V. S. E., Guo, W., Cederna, J., & Newman-Casey, P. A. (2017). Barriers to attending eye appointments among underserved adults. *Journal of Ophthalmic & Vision Research*, *12*(4), 449. doi: 10.4103/jovr.jovr_57_17

Wu, Y., Keel, S., Carneiro, V. L. A., Zhang, S., Wang, W., Liu, C., ... & He, M. (2024). Real-world application of a smartphone-based visual acuity test (WHOeyes) with automatic distance calibration. *British Journal of Ophthalmology*. https://doi.org/10.1136/bjo-2023-324913

Wu, Y. H., Granquist, C., Gage, R., Crossland, M. D., & Legge, G. E. (2020). Online survey of digital reading by adults with low vision. *Optometry and Vision Science*, *97*(4), 249-256. doi: 10.1097/OPX.0000000000001503

Xiong, Y. Z., Kwon, M., Bittner, A. K., Virgili, G., Giacomelli, G., & Legge, G. E. (2020). Relationship between acuity and contrast sensitivity: differences due to eye disease. *Investigative ophthalmology & visual science*, *61*(6), 40-40. https://doi.org/10.1167/iovs.61.6.40

Zhang, Z. T., Zhang, S. C., Huang, X. G., & Liang, L. Y. (2013). A pilot trial of the iPad tablet computer as a portable device for visual acuity testing. *Journal of telemedicine and telecare*, *19*(1), 55-59. https://doi.org/10.1177/1357633X12474964

Appendix 1 – o-VA Details

This Appendix describes the methods used to calculate the size of optotypes used in the o-VA test. Additionally, the calculations used to determine the VA thresholds for each optotype size on participants' screens are also included.

Pixel size and visual angle

A key figure needed to create stimuli for the o-VA was an estimate of pixel size. Typical pixel size estimation was done with the following process on a display with a diagonal size of 14 inches.

1. The number of pixels on the diagonal of the screen is calculated using the screen resolution.
    - Resolution: 1920 x 1080 pixels
    - Number of diagonal pixels: $\sqrt{(1080\ px)^2 + (1920\ px)^2} \approx 2203\ px$
2. Given a display size of 14 inches, pixel size is calculated with: 14 inches/2203 pixels = **0.0064 inches/pixel**
    - This calculation assumes pixels are square, not rectangular. We can check this assumption by calculating the display size based on pixel size, assuming they are square.
        i. Display height = 0.0064 inches/pixel * 1080 pixels = 6.863 inches
        ii. Display width = 0.0064 inches/pixel * 1920 pixels = 12.202 inches
        iii. Display diagonal: $\sqrt{(6.863\ inches)^2 + (12.202\ inches)^2} = 13.9996\ inches$
3. Knowing pixels are approximately 0.0064 inches/pixel, we then calculate the visual angle of a pixel at a 12-inch viewing distance.
    - First, determine half the width of a pixel: 0.0064/2 = 0.0032
    - Next, we calculate the angle one of these halves subtends
        i. $\tan(\theta) = \frac{0.0032}{12}$
        ii. $\theta = \tan^{-1}\left(\frac{0.0032}{12}\right) = 0.0153°$
    - Last, we calculate the angle of the entire pixel: 0.0153 * 2 = **0.0306° per pixel**
    - For use in measuring acuity, we translate this from degrees to arcminutes: 0.03° * 60 = **1.83 arcminutes per pixel**

Optotype threshold calculation, and range of possible optotypes

The smallest possible critical feature is 1 pixel, therefore the smallest possible optotype threshold given by: log(1.83) = 2.63. Therefore, on a 14-inch display, at a 12-inch viewing distance, the smallest possible Tumbling E stimulus (with a stroke width of 1 pixel) should be at threshold for someone with a visual acuity of 0.263 logMAR. The specific dimensions of the optotype are illustrated in Figure A1-1, below.

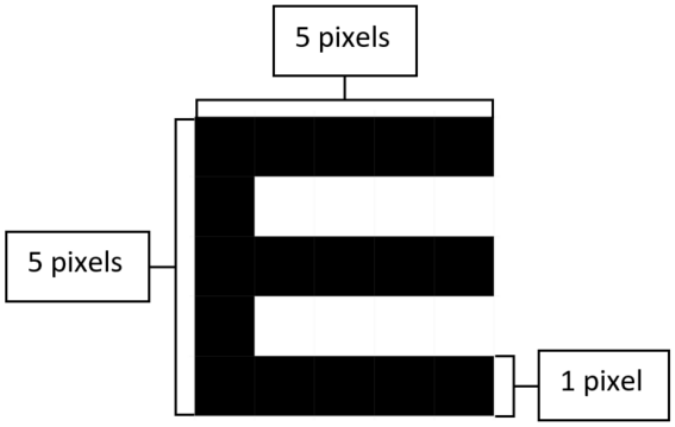


Figure A1-1: Dimensions of tumbling E stimuli.

Subsequent optotypes increase in size such that they are at threshold in 0.1 logMAR steps. Visual angle required for each .1 logMAR step was calculated by multiplying size of each subsequent angle by 1.2589, as described in Bailey & Lovie-Kitchin, 2013 (starting with (1.83*1.2589 = 2.30), then (2.30 *1.2589 = 2.90), etc). The full range of optotypes in terms of logMAR threshold, corresponding visual angle, and number of pixels used for the critical feature are listed in Table A1-1, below. The table assumes viewing on a 14-inch screen at a 12-inch viewing distance.

logMAR	Critical Feature Visual Angle (arcmin)	# Pixels Required for Critical Feature
0.26	1.830	1
0.36*	2.303	1.25
0.46*	2.900	1.58
0.56	3.651	1.99
0.66*	4.596	2.51
0.76	5.786	3.16
0.86	7.284	3.98
96	9.170	5.01
1.06	11.544	6.30
1.16	14.533	7.94
1.26	18.296	9.99
1.36	23.033	12.58
1.46	28.996	15.84
1.56	36.503	19.94
1.66	45.954	25.11
1.76	57.852	31.61

1.86	72.830	39.79
1.96	91.685	50.10
2.06	115.423	63.07
2.16	145.306	79.40
2.26	182.926	99.95
2.36	230.285	125.83
2.46	289.906	158.41
2.56	364.963	199.43

Table A1-1: Specifications for optotypes on the o-VA. A * in the logMAR column indicates that ((# pixels required) – (nearest whole number)) > |.1(# pixels required)|. This means the stimulus would be more than 10% larger or smaller than it should be (due to requiring sub-single pixel size adjustments), and was therefore not included in the test. As such, these rows are not included in future tables.

Note that because Screen 1 is 1080 pixels in height, the maximum height of a Tumbling E stroke width on that screen is 216 pixels, which corresponds to 2.6 logMAR. Therefore, the maximum optotype size possible on the 14-inch screen at a 12 in viewing distance would be 2.56 logMAR.

Optotype Threshold Calculations on Participant Displays

During the calibration process for the OVT, a square is displayed on the participant's screen. By taking a ratio of the size of the image on the participant's screen to the size of the image on the administrator screen (the Screen Scaling Factor, or SSF), it is possible to calculate how much the participant's screen is scaling the size of the images for their display. Once the SSF is determined, it can be applied to the known size of images from the administrator's screen, and subsequently the VA threshold for each optotype can be calculated. An example of this process is presented below.

1. The size of the calibration square on the administrator's screen is measured to be 3 x 3 inches.
2. The participant reports that the size of the square on their screen is 4 x 4 inches.
3. The SSF is calculated as 4/3 = 1.333.
4. The visual angles subtended by each optotype on the administrator's screen are multiplied by the SSF.
5. New VA thresholds are calculated using the new visual angles, as shown in the table below.

Administrator Screen logMAR	Administrator Screen Critical Feature Visual Angle (arcmin)	Participant Screen Critical Feature Visual Angle (arcmin)	Participant Screen logMAR
0.26	1.83	2.43	0.39
0.36*	2.30	3.06	0.49
0.46*	2.90	3.86	0.59
0.56	3.65	4.86	0.69
0.66*	4.60	6.11	0.79
0.76	5.79	7.70	0.89
0.86	7.28	9.69	0.99
96	9.17	12.20	1.09
1.06	11.54	15.35	1.19
1.16	14.53	19.33	1.29
1.26	18.30	24.33	1.39
1.36	23.03	30.63	1.49
1.46	29.00	38.57	1.59
1.56	36.50	48.55	1.69
1.66	45.95	61.12	1.79
1.76	57.85	76.94	1.89
1.86	72.83	96.86	1.99
1.96	91.69	121.94	2.09
2.06	115.42	153.51	2.19
2.16	145.31	193.26	2.29
2.26	182.93	243.29	2.39
2.36	230.29	306.28	2.49
2.46	289.91	385.58	2.59
2.56	364.96	485.40	2.69

Table A1-2: Size of optotypes and corresponding logMAR thresholds on a screen which increases image sizes by 1.333 fold.

Note that in the example in the table above, the range of the test shifts from 0.26 – 2.56 logMAR to 0.39 – 2.69 logMAR. The direction of this shift, toward poorer vision, is an intuitive one, once it is considered that the optotypes were scaled up to be larger on the participant's screen. Also note that all the above calculations can be adjusted for different viewing distances. This is done by changing the viewing distance in the original calculation of pixel size, then following all the same steps till this point.

Appendix 2 – o-CS Details

This Appendix describes the process of creating the dither patterns used in the o-CS test, from initial conception, to photometric testing, and examples of completed dither patterns. Luminance measurements were taken using a Minolta CS-100 Colorimeter photometer (Konica Minolta Sensing Americas, Inc., Ramsey, NJ). For all measurements, a tripod was used to point the photometer at the computer screen the pattern was displayed on, such that the photometer was angled orthogonally to the screen. Measurements were taken in a darkened room with no sources of illumination aside from the screen.

Calculation of Proportion of White and Black Pixels for Dither Patterns

Before creating the dither patterns, we first quantified the ratio of light emitted from white compared to black pixels, and determined the relationship between screen luminance and the ratio of white to black pixels on screen. We hypothesized there would be a 1:1, linear relationship between screen luminance and the proportion of white pixels on screen (expecting black pixels to emit very little light), and expected relationship this would be independent of the proportion of pixels that were white.

Nine preliminary dither patterns (not for use in the actual OVT) were created to evaluate that hypothesis. The patterns were 10 by 10 grids of white and black pixels, henceforth called “tiles”, with percentages of white pixels ranging from 10% to 90%. The luminance of each tile was measured by creating an image that was equal in size to a computer monitor, and filled entirely with copies of the tile in question. Screens comprised entirely of black pixels and entirely of white pixels were also measured to determine a minimum and maximum screen luminance. This initial process was conducted on two displays: one connected to a computer running on Windows 10 OS (Dell Latitude laptop), the other connected to a computer running on MacOS 12 (Macbook Air).

As each display had a different maximum luminance, we converted luminance measurements for each dither pattern into a percentage of maximum luminance (the luminance of the dither pattern divided by the luminance of an all-white pixel screen). We found that for the Windows OS display, our prediction of a linear relationship between luminance and proportion of white pixels was correct. Each 10% increase in the number of white pixels corresponded to a 10% increase in percent of maximum luminance. This relationship was also tested at all screen brightness levels. White the actual luminance emitted was affected by screen brightness, the relationship between the number of pixels on screen and the maximum possible luminance remained the same.

However, we found a different pattern for the MacOS. There, we observed a curve with a shallower slope at low percentages of white pixels, and an increasingly steep slope

at higher percentages. To determine if this was an issue with that specific screen, the same procedure was carried out with two additional displays connected to computers also running macOS (Macbook Pro and iMac). The same pattern was observed in the other screens as well, suggesting this effect was due to a form of brightness adjustment in the macOS. The data from these measurements were plotted (with an interpolated line between data points), with the percent of the dither patterns being black pixels on the x-axis, and their percentage of max luminance on the y-axis, shown in Figure A2-1. Note that the data points for the macOS measurements are the mean values of all three screens. These curves were called the "Expected Luminance Curves," as they were used to inform what we would expect the luminance of a dither pattern to be with a specified proportion of white to black pixels. These expected curves for macOS were also tested at all brightness levels of the screen, and the curve remained the same regardless of screen brightness.

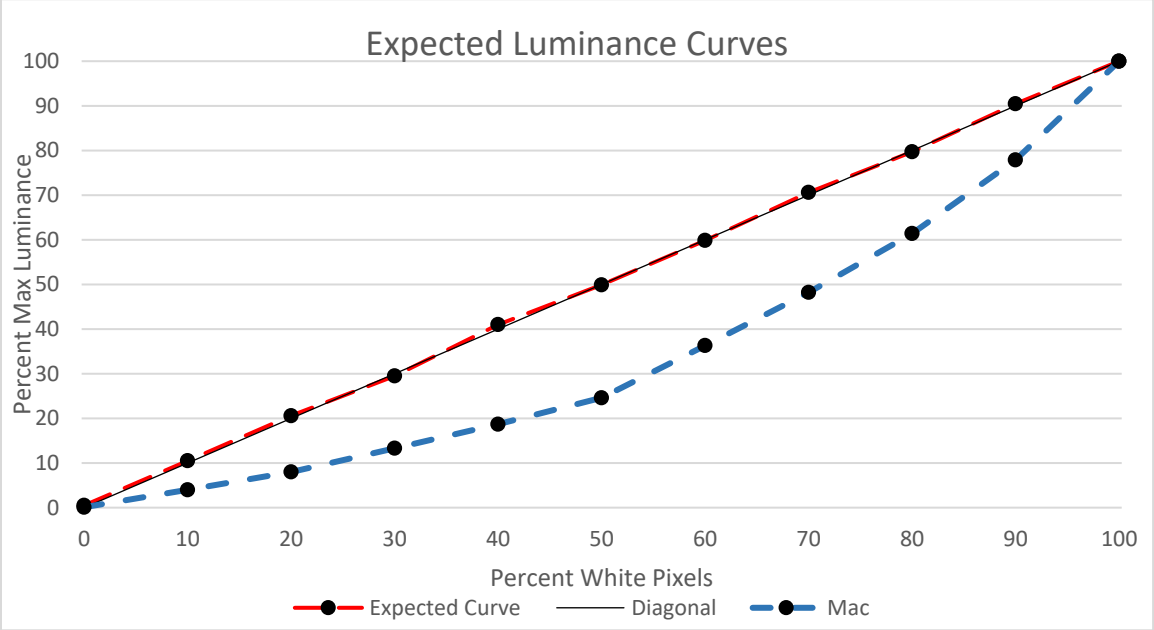


Figure A2-1: Expected vs measured luminance values at different proportions of white and black pixels on a MacOS display.

The white to black pixel ratios for the dither patterns that would be used for the actual test were then calculated. This was first done for the Windows OS patterns, using the following equation:

$$\log_{10}(\frac{1}{C}) = X$$

Where X is the desired logCS, and C is Michelson contrast, calculated with:

$$C = \frac{(L_{max} - L_{min})}{(L_{max} + L_{min})}$$

Where L_{max} is the brightest part of the image (the white background in standard polarity, or the E optotype in reverse polarity), and L_{min} is the darkest part of the image (the E optotype in standard polarity, or the background in reverse polarity). In these equations, we input whatever logCS value is needed, and solve for L_{min}. Importantly, we aimed to determine a percentage of pixels that would need to be white for each dither pattern, which corresponded to a percentage of L_{max}, rather than any specific luminance value, so L_{max} was simply set to 1. This way, when L_{min} is solved for, the result is a number from 0 to 1, and when multiplied by 100, equals the percentage of white pixels necessary for a dither pattern at the given logCS (for Windows OS). An example of the calculation process is shown below, calculating a percentage of pixels that need to be white in a dither pattern to create a contrast with a white background that corresponds to a contrast threshold of logCS = 1.

$$\log_{10}(\frac{1}{C}) = 1$$

$$10^1 = 10 = \left(\frac{1}{c}\right)$$

$$10C = 1$$

$$C = \frac{1}{10} = .1$$

$$.1 = \frac{1 - L_{min}}{1 + L_{min}}$$

$$.1 + .1L_{min} = 1 - L_{min}$$

$$.1L_{min} = .9 - L_{min}$$

$$1.1L_{min} = .9$$

$$L_{min} = .818$$

$$Percentage\ of\ white\ pixels = 81.8\%$$

Therefore, for an observer with logCS = 1, a dither pattern would be at the contrast threshold (when against a white background) when it contains 81.8% white pixels. These

percentages were calculated for each logCS step from 0.1 to 2, in 0.1 log unit steps for Windows OS screens. Note that the stimuli (i.e. dither patterns) being created were 10x10 grids of pixels, so percentages were rounded to the nearest whole percent (in this case, 82 pixels were used). The number of white pixels used in each dither pattern is included in Table A2-1 below.

The first round of percentages were equal to the number of white pixels needed for each 10 x 10 dither pattern tile for Windows OS displays, but macOS displays required an additional step due to their percentage of white pixels to luminance nonlinearity. The expected curve for macOS was used here; the value on the x-axis at which the curve intercepted the value given by the percentage calculation was used as the indicator for what the proportion of white pixels should be. For instance, for logCS = 0.5, L_{min} is equal to 0.52. The point at which the macOS expected curve reaches a y-value of 52 is at x = .707, which rounds up to 71% white pixels. This process was repeated to calculate the percentages of white pixels for each logCS step from 0.1 to 2, in 0.1 log unit steps for MacOS screens.

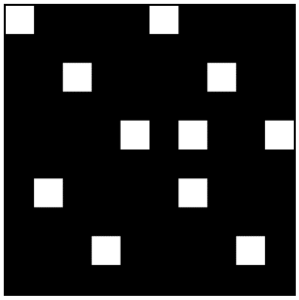
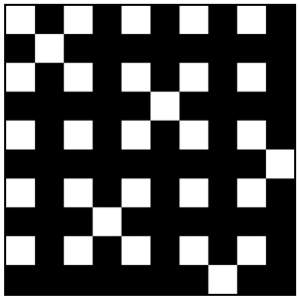
logCS	# White Pixels for Windows Dither Patterns	# White Pixels for macOS Dither Patterns
0.1	11	30
0.2	23	47
0.3	33	57
0.4	44	64
0.5	52	71
0.6	60	75
0.7	67	80
0.8	73	83
0.9	78	86
1.0	82	89
1.1	85	91
1.2	88	93
1.3	90	94
1.4	92	95
1.5	94	96
1.6	95	193/200
1.7	96	97
1.8	97	195/200
1.9	98	391/400
2.0	99	98

Table A2-1: LogCS progression and corresponding number of white pixels (out of 100) in each tile.

For some log steps in the macOS dither patterns, particularly at and above 1.6 logCS, a larger grid needed to be created. This was because the required luminance steps were less than 0.5%, which was not possible when working with only 100 pixels, as adding a single white pixel would increase the luminance by 1%, and thus skip a step. For example, the number of white pixels required for a stimulus to be at threshold for 1.6 logCS is 96.5, which would typically round up to 97, but 1.7 logCS requires 97 white pixels. Similarly, 1.8 requires 97.5, which would round to 98, but 1.9 requires 97.75, and 2.0 also requires 98. To more accurately create these luminance values, grids of 10x20 pixels were used for 1.6 and 1.8, and a grid of 20x20 was used for 1.9. Table A2-1 provides more specific information on these tiles, as these entries are fractions in which the numerator is the number of white pixels, and the denominator is the total number of pixels.

Production of Dither Pattern Tiles and Implementation for E Optotypes

A set of dither pattern tiles were created for both Windows OS and macOS, according to the proportions of white and black pixels in Table A2-1. This was done using the design software Figma (Version 116.13.3), which allows for pixel-by-pixel construction of image files. Each dither pattern was started with a 10 x 10 grid of black pixels (or a larger grid for some macOS patterns), and was then individually edited by hand to have the appropriate proportion of white to black pixels. Each pattern tile was designed to create as uniform a texture as possible, and to avoid obvious repeating patterns, both within a single tile, and when multiple tiles were placed adjacent to each other in a larger image. Examples of the completed tiles are shown below.

	Operating System	
LogCS	Windows	macOS
0.1	11 White Pixels 	30 White Pixels 

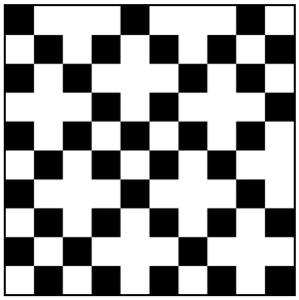
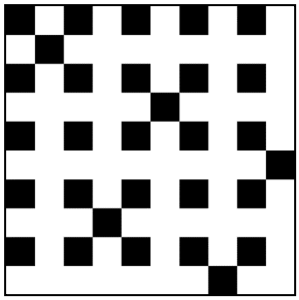
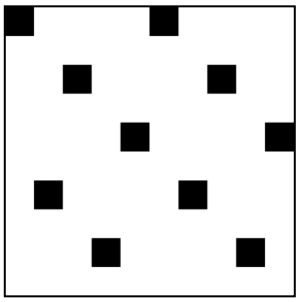
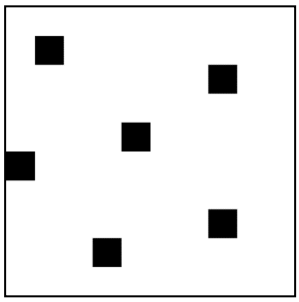
0.6	60 White Pixels 	70 White Pixels 
1.3	90 White Pixels 	94 White Pixels 

Table A2-2: Examples of tiles created for each OS at various logCS thresholds.

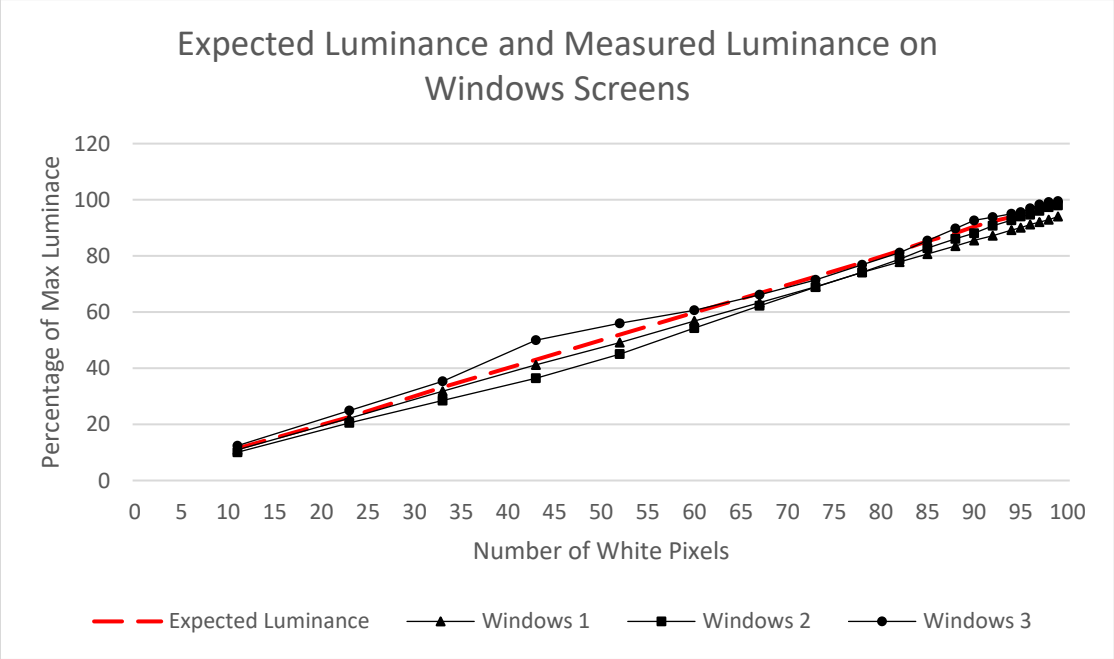
Photometric Validation of Dither Patterns

The luminance emitted by each dither pattern was measured using the same photometer and methods used to create the expected curves. The following tables display the results of luminance measurements for the dither pattern stimuli, taken on three separate screens for both Windows and Mac stimuli sets. The luminance data are then converted to the percentage of max luminance for each screen, and are then plotted below. Note that rather than starting at 100% white pixels and ending at 0%, the table only includes values that were used to create stimuli for the test.

Windows – Absolute Values (cd/m^2)

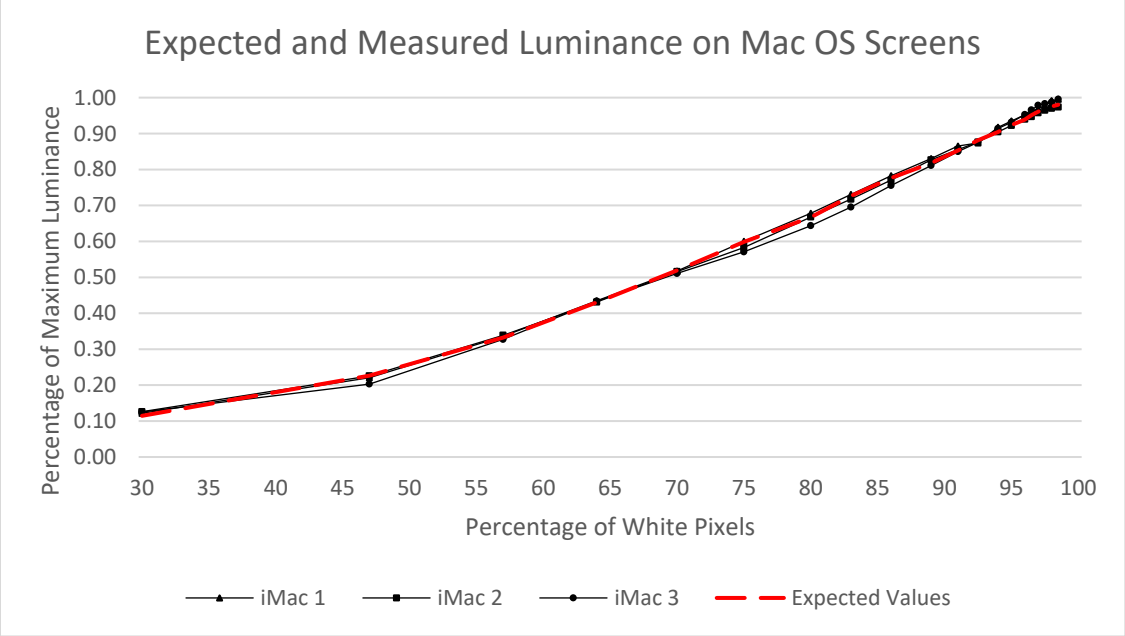
logCS (% White Pixels)	Windows 1		Windows 2		Windows 3	
	Measured	Expected	Measured	Expected	Measured	Expected
0.1 (11)	38.5	40.3	15.2	17.3	42.2	39.0
0.2 (23)	78	79.6	31	34.2	84.7	76.9
0.3 (33)	112	117	43	50.2	120	113

0.4 (43)	145	152	55	65.0	170	146
0.5 (52)	173	183	68	78.4	190	177
0.6 (60)	200	211	82	90.4	206	203
0.7 (67)	223	235	94	101	225	227
0.8 (73)	243	256	104	110	243	247
0.9 (78)	261	273	112	117	261	264
1.0 (82)	274	288	119	124	276	278
1.1 (85)	284	300	125	129	290	290
1.2 (88)	294	310	130	133	305	300
1.3 (90)	301	318	133	137	315	308
1.4 (92)	307	325	137	139	319	314
1.5 (94)	314	330	140	142	323	319
1.6 (95)	317	335	142	144	325	323
1.7 (96)	321	338	143	145	329	327
1.8 (97)	324	341	145	146	334	329
1.9 (98)	327	343	147	147	337	332
2.0 (99)	331	345	148	148	338	333
White	352		151		340	



iOS – Absolute Values (cd/m²)

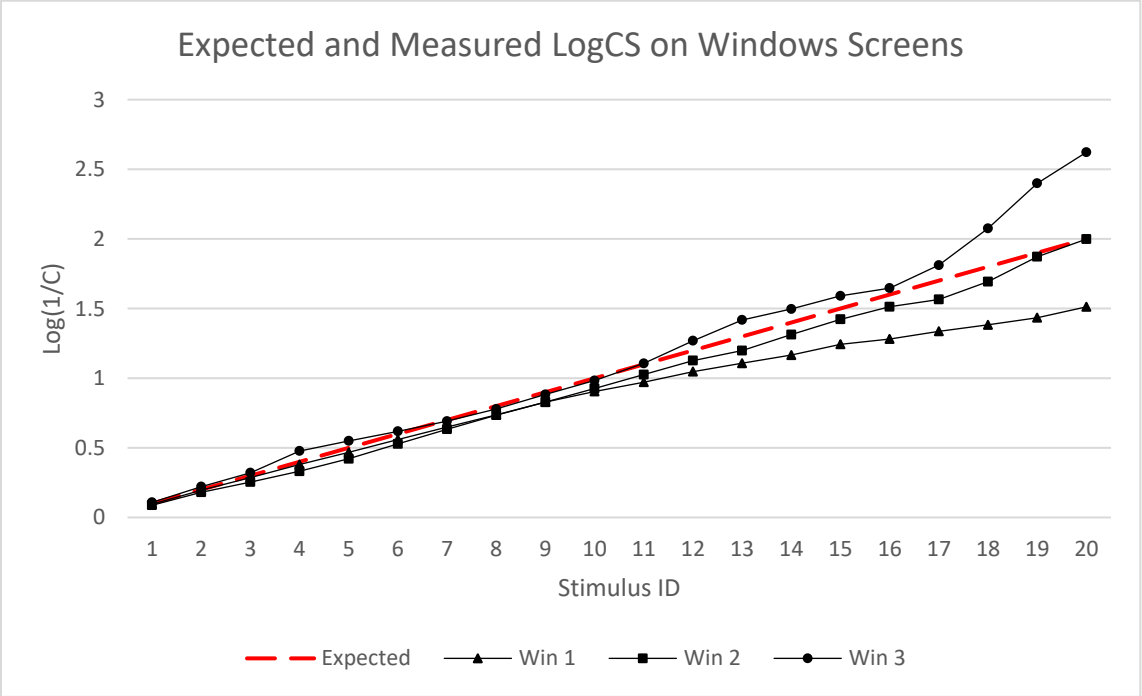
logCS (% White Pixels)	Mac OS 1		Mac OS 2		Mac OS 3	
	Measured	Expected	Measured	Expected	Measured	Expected
0.1 (30)	28	26.4	35.7	32.4	29.2	26.7
0.2 (47)	50.8	52.0	63.9	64.0	47.3	52.7
0.3 (57)	77.3	76.4	95.9	94.0	76.3	77.4
0.4 (64)	100	99.0	122	122	101	100
0.5 (70)	119	119	146	147	119	121
0.6 (75)	138	138	165	169	133	139
0.7 (80)	156	153	189	189	150	155
0.8 (83)	168	167	203	206	162	169
0.9 (86)	180	179	218	220	176	181
1.0 (89)	191	188	234	232	189	191
1.1 (91)	199	196	242	241	198	199
1.2 (92.5)	201	203	248	249	204	205
1.3 (94)	211	208	256	256	213	211
1.4 (95)	215	212	261	261	217	215
1.5 (96)	219	216	266	266	222	219
1.6 (96.5)	221	219	268	269	225	222
1.7 (97)	223	221	271	272	228	224
1.8 (97.5)	225	223	273	274	229	226
1.9 (97.75)	228	224	274.5	276	230	227
2.0 (98)	229	225	275.5	277	232	228
White	230		283		233	



The tables here contain the luminance data from the previous tables, but converted to their corresponding logCS values. Below each table, the logCS values from each screen are plotted, along with a diagonal line to indicate ideal logCS values.

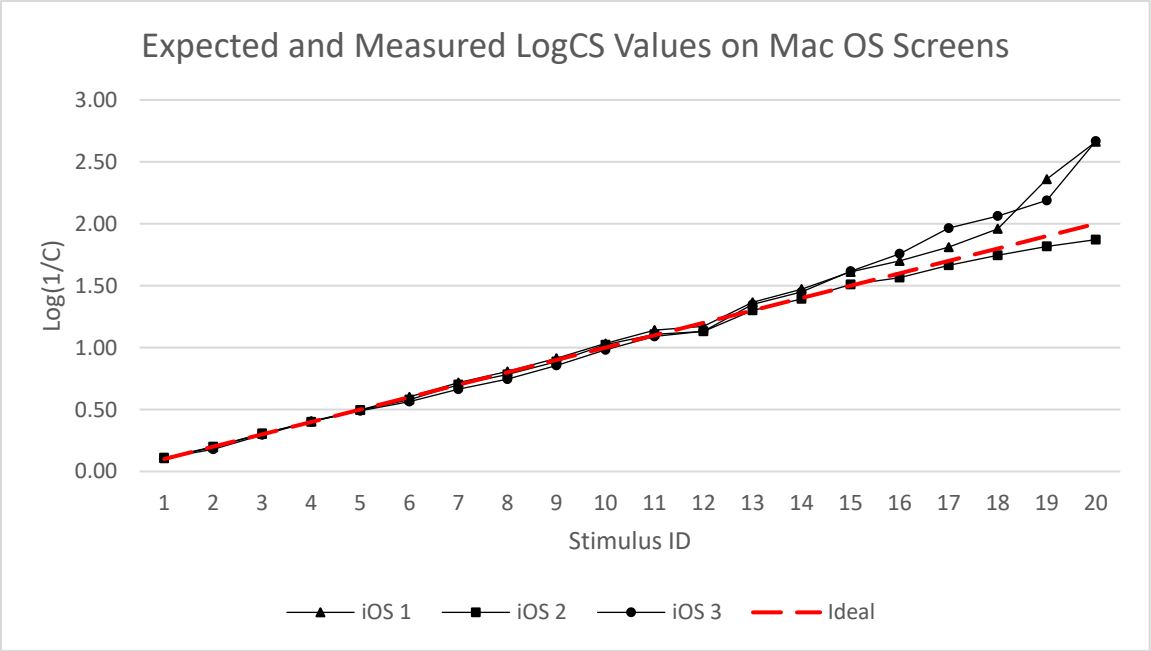
Windows

logCS (% White Pixels)	Measured			Expected	Mean	Mean Deviation
	Windows 1	Windows 2	Windows 3			
0.1 (11)	0.095	0.088	0.108	0.100	0.097	0.008
0.2 (23)	0.196	0.181	0.221	0.200	0.199	0.015
0.3 (33)	0.286	0.254	0.321	0.300	0.287	0.027
0.4 (43)	0.380	0.332	0.477	0.400	0.396	0.055
0.5 (52)	0.467	0.421	0.550	0.500	0.480	0.054
0.6 (60)	0.560	0.529	0.619	0.600	0.569	0.043
0.7 (67)	0.649	0.633	0.691	0.700	0.658	0.042
0.8 (73)	0.737	0.734	0.779	0.800	0.750	0.050
0.9 (78)	0.828	0.829	0.883	0.900	0.847	0.053
1.0 (82)	0.904	0.926	0.983	1.000	0.938	0.062
1.1 (85)	0.971	1.026	1.106	1.100	1.034	0.070
1.2 (88)	1.047	1.126	1.269	1.200	1.147	0.098
1.3 (90)	1.107	1.198	1.418	1.300	1.241	0.138
1.4 (92)	1.166	1.313	1.497	1.400	1.325	0.139
1.5 (94)	1.244	1.423	1.591	1.500	1.419	0.142
1.6 (95)	1.281	1.513	1.647	1.600	1.480	0.151
1.7 (96)	1.337	1.565	1.812	1.700	1.571	0.203
1.8 (97)	1.383	1.693	2.076	1.800	1.717	0.267
1.9 (98)	1.434	1.872	2.399	1.900	1.902	0.331
2.0 (99)	1.512	1.999	2.622	2.000	2.044	0.371



iOS

logCS (% White Pixels)	Measured			Expected	Mean	Mean Deviation
	Mac OS 1	Mac OS 2	Mac OS 3			
0.1 (30)	0.106	0.110	0.109	0.1	0.109	0.009
0.2 (47)	0.195	0.200	0.179	0.2	0.191	0.009
0.3 (57)	0.304	0.306	0.295	0.3	0.302	0.005
0.4 (64)	0.405	0.401	0.403	0.4	0.403	0.003
0.5 (70)	0.498	0.496	0.490	0.5	0.494	0.006
0.6 (75)	0.602	0.579	0.563	0.6	0.582	0.020
0.7 (80)	0.717	0.701	0.664	0.7	0.694	0.018
0.8 (83)	0.807	0.784	0.745	0.8	0.779	0.026
0.9 (86)	0.914	0.887	0.856	0.9	0.886	0.024
1.0 (89)	1.033	1.023	0.982	1	1.013	0.025
1.1 (91)	1.141	1.107	1.090	1.1	1.113	0.019
1.2 (92.5)	1.172	1.131	1.132	1.2	1.145	0.055
1.3 (94)	1.366	1.300	1.348	1.3	1.338	0.038
1.4 (95)	1.472	1.393	1.449	1.4	1.438	0.043
1.5 (96)	1.611	1.509	1.617	1.5	1.579	0.079
1.6 (96.5)	1.700	1.565	1.758	1.6	1.674	0.098
1.7 (97)	1.811	1.664	1.965	1.7	1.813	0.137
1.8 (97.5)	1.959	1.745	2.063	1.8	1.922	0.159
1.9 (97.75)	2.360	1.817	2.188	1.9	2.122	0.277
2.0 (98)	2.662	1.872	2.667	2	2.400	0.486

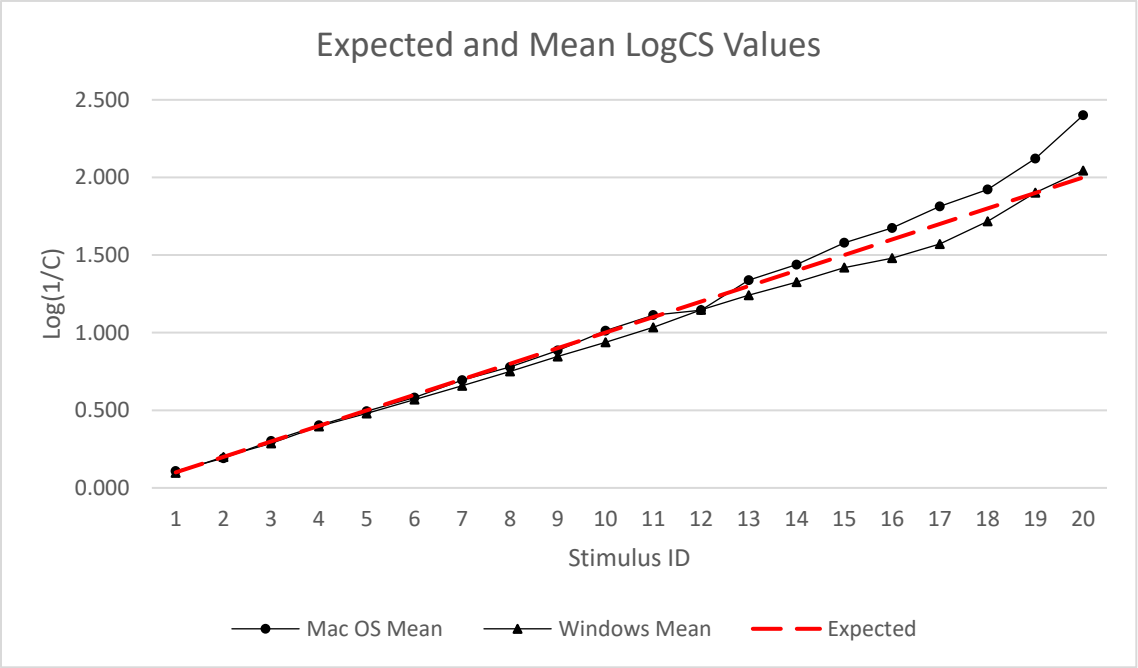


It is notable that for both Windows and macOS displays, logCS values were both more in line with what was intended and also more reliable across screens at lower luminance values. This can be attributed to the fact that at lower contrast values, even very small differences in luminance values lead to a very large difference in logCS. Thus, even though the different screens had very similar luminance outputs, at this range of stimuli, the screens were displaying different logCS values. This suggests that the on-screen testing will be more reliable at poorer logCS ranges.

To summarize these data, this table lists the mean logCS values across all three screens for both display types, along with the corresponding intended logCS value. A plot is also included for a visual reference for these data.

Expected	Windows	Mac OS
0.1	0.097	0.109
0.2	0.199	0.191
0.3	0.287	0.302
0.4	0.396	0.403
0.5	0.480	0.494
0.6	0.569	0.582
0.7	0.658	0.694
0.8	0.750	0.779
0.9	0.847	0.886
1	0.938	1.013
1.1	1.034	1.113
1.2	1.147	1.145
1.3	1.241	1.338

1.4	1.325	1.438
1.5	1.419	1.579
1.6	1.480	1.674
1.7	1.571	1.813
1.8	1.717	1.922
1.9	1.902	2.122
2	2.044	2.400



Appendix 3 – o-VF Participant Instructions

The o-VF test was designed to estimate the size of the visual field, in terms of degrees of visual angle horizontally and vertically. Specifically, we aimed to ascertain if we could gather a rough estimate of the horizontal and vertical field, in terms of degrees of visual angle remaining. In the o-VF, participants are presented with either a vertical or horizontal black bar against a white background. The bar extends across the entire length of the screen, either from top to bottom or from left to right. Participants are instructed to determine if they are able to see both edges of the black bar simultaneously, in one glance, without moving their eyes. The participants were given instructions which included a visualization exercise to further explain this idea, which are stated below.

"We will now test your visual field.

As a primer to explain the test before we begin, imagine you are sitting in a movie theater, looking at the screen. If you were sitting in the front row, you would likely not be able to see both edges of the screen at the same time – the screen would be too big. If you moved to the middle of the theater, there would be a seat in which you could sit where the screen would fill your whole field of view, but you would not be able to see things past the edge of the screen. If you were to sit in the back row, you would likely be able to see both edges of the screen at once, as well as the walls on either side of it.

During this test, we will use a similar principle to explore this aspect of your vision, determining what size images will fill your field of view. During this test, a black bar will appear against a white background.

Your task is to report whether you can see both edges of the bar at once. For the first round, the bar will first be oriented vertically, and you will report if you can see both the left and right edges at the same time. For the second round, it will be oriented horizontally, and you will report if you can see both the top and bottom edges at the same time. The bar will steadily get larger, until we get to the maximum size, then it will get smaller again.

You can move your eyes to find the edges, but after you find them, you need to report whether you can see both edges of the bar simultaneously, without moving your eyes back and forth. Try not to move your head closer to or farther away from the screen. Like the previous test, I will give you a verbal signal, saying "now," when the bar appears. Do you have any questions?"