New Technologies in Paediatric Acuity Assessment

Iain Alexander Thomas Livingstone

BSc (Hons) Biomedical Sciences, Pharmacology
MB ChB
FRCOphth

Submitted in fulfilment of the requirements for the Degree of Doctor of Medicine

College of Medical, Veterinary and Life Sciences
University of Glasgow

November 2019

Declaration

I declare that, except where explicit reference is made to the contribution of others, that this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Signature:

Printed name:

Iain Livingstone

Details of work and collaboration

Chapter 2

IL was funded in part by the Queen Elizabeth Diamond Jubilee Trust.
URL: http://jubileetribute.org. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conceived and designed the experiments:

I Livingstone, CM Tarbert, ME Giardini, R Hamilton

Performed the experiments:

I Livingstone, CM Tarbert

Analysed the data:

I Livingstone, CM Tarbert, ME Giardini, R Hamilton, A Bastawrous, D Middleton

Contributed reagents/materials/analysis tools:

I Livingstone, CM Tarbert, ME Giardini, R Hamilton

Wrote the paper:

I Livingstone, CM Tarbert, ME Giardini, R Hamilton, D Middleton

Chapter 3

Conceived and designed the experiments:

I Livingstone, ME Giardini, R Hamilton, G Vivas

Performed the experiments:

I Livingstone, G Vivas, A Cheema

Analysed the data:

I Livingstone, ME Giardini, R Hamilton, G Vivas

Contributed reagents/materials/analysis tools:

I Livingstone, ME Giardini, R Hamilton, G Vivas

Wrote the paper:

I Livingstone, ME Giardini, R Hamilton, A Cheema, G Vivas

Chapter 4

Conceived and designed the experiments:

I Livingstone, R Hamilton

Performed the experiments:

I Livingstone, A Lok

Analysed the data:

I Livingstone

Contributed reagents/materials/analysis tools:

I Livingstone, R Hamilton

Wrote the paper:

I Livingstone

Chapter 5

Conceived and designed the experiments:

I Livingstone, R Hamilton

Performed the experiments:

I Livingstone, L Butler, E Misanjo, A Lok, S Delfin

Analysed the data:

I Livingstone, L Butler, R Hamilton

Contributed reagents/materials/analysis tools:

I Livingstone, R Hamilton, L Butler, D Middleton

Wrote the paper:

I Livingstone, L Butler, R Hamilton

The research within Chapter 5 was supported by Fiona's Eye Fund (study in Malawi) and by the Queen Elizabeth Diamond Jubilee Trust for equipment and research time (IL) (study in United Kingdom).

Peekaboo Vision is a CE marked medical device. The legal manufacturer for Peekaboo Vision is Scottish Health Innovation Ltd., who handle intellectual property on behalf of the National Health Service in Scotland.

None of the authors have a commercial relationship with Scottish Health Innovations Ltd

Abbreviations

BS:	British Standards
DR:	Diabetic Retinopathy
BEO:	Both Eyes Open
CAC:	Cardiff Acuity Cards
CG:	Chessboard Grating
cd/m^2:	Candela per meter squared
CR:	Coefficient of Repeatability
CRT:	Cathode Ray Tube
CE:	Conformité Européene (European Conformity)
ETDRS:	Early Treatment Diabetic Retinopathy Study
ES:	Engagement Score
FCPL:	Forced choice preferential looking
FDA:	Food and Drug Administration
FG:	Finest Grating
HG:	Horizontal Grating
HTML5:	Hypertext Markup Language (5)
ICO:	International Council of Ophthalmology
JND:	Just Noticeable Difference
KACI:	Keeler Acuity Cards for Infants
PV:	Peekaboo Vision
PVb1:	Peekaboo Vision Prototype-Build 1
PVi:	Peekaboo Vision for iOS
K:	Kelvin
L_{black}:	Black Luminance
L_{bkg} :	Background Luminance
LE:	Left Eye
LED:	Light Emiting Diode
$L_{grating}$:	Grating Luminance
L_{letter}:	Letter Luminance
LoA:	Limits of Agreement
LogMAR:	Logarithm of Mean Angle of Resolution
LP:	Lea Paddles
Lux:	Illuminance / luminous emittance
L_{white}:	White Luminance
Lx:	Lumens
MC:	Michelson contrast
NM:	Not Measured
OCT:	Optical Coherence Tomography
OpenCV:	Open Computer Vision Libraries
P:	Probability value
RE:	Right Eye
RGB:	Reg/Green/Blue
ROP:	Retinopathy of Prematurity
SFG:	Second Finest Grating
TAC:	Teller Acuity Cards
VG:	Vertical Grating
W:	Wilcoxon statistic
WC:	Weber contrast

Abstract

The present research has evaluated the utility of the computer tablet as a means to test vision in an infant population. The following points summarise the main findings:

Regards the physical properties of mobile tablet computer with reference to National and International Standards for Chart Design:

- Photometric standards of luminance, contrast, and luminance uniformity are met more effectively by the range of iPads under test than by dedicated ETDRS (Early Treatment Diabetic Retinopathy Study) charts in active clinical use in a tertiary referral unit. The study met its aim of documenting the suitability of this mobile technology regards high contrast acuity standards, providing a practical guide to health care professionals working within eye care.

These standards were then, where possible, extended to card-based infant vision tests.

- There are intrinsic advantages to digital platforms regards achieving a mean average luminance "grey" background relative to the black and white values of composite foreground gratings. There is marked heterogeneity across traditional card-based platforms regards luminance and contrast measurements relating to black, white and grey components.

These observations informed the design of a prototype build of a computer-tablet based infant acuity test, peekaboo vision (PVb1). This build was evaluated in a blurred adult cohort.

- PVb1 performed well across a range of artificially degraded acuities, with observed potential benefits regards test-retest repeatability.

PVb1 was then evaluated in an infant population in rural Africa in a pilot study (Study

1). A subsequent formal build (PVi) was evaluated in a UK setting with a similar methodology (Study 2), comparing with Keeler Acuity Cards for Infants (KACI) as the reference standard.

Across Studies 1 and 2, the mean difference between reference standard and digital version was modest (-0.03 to 0.01), with notable differences in upper and lower limits of agreement in favour of the digital platform (exhibiting narrower LoA). Peekaboo Vision evidenced improved repeatability than KACI: coefficients of repeatability were 0.27 for Peekaboo Vision versus 0.37 for Keeler cards in study 1, and 0.32 for Peekaboo Vision versus 0.42 for Keeler cards in study 2. The mean time-to-test was over 1 minute shorter (by 26%) for Peekaboo Vision than for Keeler cards ($p= 0.0021$).

Publications associated with this research

The present thesis is formed from 3 core published papers:

1. Livingstone I A. T, Tarbert CM, Giardini ME, Bastawrous A, Middleton D, Hamilton R. Photometric Compliance of Tablet Screens and Retro-Illuminated Acuity Charts As Visual Acuity Measurement Devices. PLOS ONE. 2016 Mar 22;11(3)

This paper comprises the crux of Chapter 2

Livingstone I A T, Lok ASL, Tarbert C. New mobile technologies and visual acuity. Conf Proc IEEE Eng Med Biol Soc. 2014;2014:2189–92.

This paper comprises the crux of Chapter 3

Livingstone I, Butler L, Misanjo E, Lok A, Middleton D, Wilson JW, et al. Testing Pediatric Acuity With an iPad: Validation of "Peekaboo Vision" in Malawi and the UK. Trans Vis Sci Tech. 2019 Jan 2;8(1):8–8.

This paper comprises the crux of Chapter 5

Correspondence regards permission to use published content is included in Appendix B

Published abstracts from the present work:

Butler LA, Misanjo E, Middleton D, Livingstone I, Kayange P. Evaluation of a Novel Digital Infant Acuity Test. Invest Ophthalmol Vis Sci. 2015 Jun 11;56(7):3196–3196.

Published papers in parallel work on a related theme:

Bastawrous A, Hennig B, Livingstone I. mHealth Possibilities in a Changing World. Distribution of Global Cell Phone Subscriptions. Journal of Mobile Technology in Medicine. 2013 May 1;2(1S):22–5

Bastawrous A, Rono HK, Livingstone I, et al. Development and validation of a smartphone-based visual acuity test (peek acuity) for clinical practice and community-based fieldwork. JAMA Ophthalmol

Bolster NM, Giardini ME, Livingstone IA, Bastawrous A. How the smartphone is driving the eye-health imaging revolution. Expert Review of Ophthalmology. 2014 Dec 1;9(6):475–85.

Acknowledgements

I would like to thank Dr Ruth Hamilton for her continued interest, guidance, clarity, patience and meticulous documentation, which made wading through the excel docs so much less painful.

Thanks to contributing funders: The Queen Elizabeth Diamond Jubilee Trust, National Eye Research Council, Rosetree's Trust and especially Fiona's Eye Fund, who have consistently provided support to trainees across Glasgow furthering research towards better eye care in low- and middle-income settings.

I would also like to thank Mr Alan S Lok, Dr Laura Butler and Dr Silfia Deflin for their help with data collection during the paediatric acuity trials.

Special thanks to Dr Duncan Middleton and Dr Sandy Weir from the Medical Devices Unit, for their expertise and attention to detail in the development of Peekaboo Vision for iOS.

Thanks also to Guillermo Vivas for his help with the experiments in Chapter 3, leading the way with his work on interrogating the physical properties of the index and reference tests.

Having my dear friend and colleague, Dr Mario Giardini. His contribution to this line of investigation has been transformative. Saluti!

Thanks to my long-suffering brother, Stuart, who I am presently keeping late at work as he is going to print this thesis. Thanks Donk.

Thanks to Fee and Vix for your consistent support.

Thanks finally to my family. To Lucy, Daisy and Angus for staying in good humour while I miss too many Saturdays writing this thesis and derail our family holiday, as always chasing at the 11th hour.

Table of contents

List of figures

List of tables

Chapter 1: Introduction

1.1 Background and aims

1.1.1 Infant Vision and Amblyopia

It is estimated that in almost half of the children who are blind today, the underlying cause could have been prevented, or the eye condition treated to preserve vision or restore sight[1]. 63% of childhood visual impairment worldwide is due to refractive error (poor vision correctable with glasses) and treatment during the phase of rapid development helps avoid permanent impairment[2,3].

Amblyopia is a neurodevelopmental disorder arising due to interruption/alteration of usual processes during the early sensitive period of visual development. The most common cause is *refractive* amblyopia: visual blur due to the image defocus secondary to refractive error. Other causes include misalignment of the eyes (strabismic amblyopia), and structural disorders causing obscuration of an image (form-deprivation amblyopia), which may arise from media opacity (such as lens, corneal or vitreous opacity), or ptosis. Amblyopia may be unilateral or bilateral, and is typically associated with defective stereoacuity (depth of vision)[4].

Amblyopia represents the most common visual disorder across all age groups, with a cumulative incidence is estimated between 2% - 4% up to 15 years old[3]. Simple patching and spectacle correction represent the most common intervention.

Typical practice is to institute treatment during the *critical period*, thought to span from early infancy to around 8 years of age, to maximise visual potential and restore a more normal trajectory of visual development[5].

It is accepted that preschool vision screening can help in the early detection of reduced

visual acuity[6]. The Royal College of Ophthalmologists guidelines state that a visual assessment by an orthoptist should be carried out on all children between the ages of four and five years [7]. This screening is aimed at detecting unsuspected visual impairment in one or both eyes. Children who achieve less than 0.2 LogMAR (6/9.5 Snellen) in either eye, despite good co-operation should be referred. However, it has been reported that there is a deficiency in vision screening specialists, falling short of such a stipulation, as investigated by Consumer company Which? (2011)[8]. It was found that 1 in 5 primary care trusts in England were providing inadequate levels of visual screening for pre-school children between the ages of four and five years.

The earlier the phase of life, the greater the potential visual problems to have a deleterious lifelong impact [9]. Earlier interventions are recognised to have maximal benefit [10], with children reported to have a 4-fold increase in remaining amblyopic when screened only at 37 months, versus repeated screens between 8 – 37 months [11].

Early treatment of poor vision in a child is hence of vital importance, as uncorrected poor vision in one or both eyes precludes the normal development of the visual brain, resulting in amblyopia. The 'critical period' of visual maturation is likely to vary and overlap for each visual function, with anatomically *higher* levels denoting a *later* critical period [12].

The crucial first step is detection. A recent convergence of technologies combines high-resolution touchscreen devices with advanced computing power, and connectivity to an electronic patient record, bringing potential to improve upon existing standards with low-cost portable solutions.

Paradoxically, the cost of a computer tablet is significantly lower than current card-

based standardised infant acuity tests. The touchscreen high-resolution display brings the potential for a robust vision test in the form of a computer game for children. A standardised stair-casing paradigm with automatic reporting could reduce the necessity for highly trained specialists to conduct testing, and consequently extend the reach of visual screening programs via with the ubiquitous online app markets.

A major motivation from the present study relates to the paucity of development of infant vision tests, with the standards being used clinically unchanged over the past 3 decades[13–16]. This is in stark contrast to the dramatic pace of change regards diagnostic modalities in other fields of ophthalmic diagnostics, such as the Deep Learning techniques being applied to various computer vision modalities to allow retinal imaging/OCT (Optical Coherence Tomography) interpretation towards screening in Retinopathy of Prematurity (ROP)[17] and Diabetic Retinopathy (DR)[18].

1.1.2 Acuity in Infants and Young Children: Conventional Testing Methods

The perfect vision test for a young child would provide indices of each visual function, profiling optic nerve function and providing information specific to cerebral visual impairment, including stereopsis. The perfect amblyopia screening test would spot at risk children as *early as possible* to enable connection with a proven intervention. The present thesis focuses squarely on high contrast acuity, which represents only one facet of visual function, and only one facet of the gamut of tests that would be used in the comprehensive visual assessment of a child.

Contemporary high contrast infant acuity tests share a singular core: forced choice preferential looking (FCPL) with high- spatial frequency targets set on an isoluminant backdrop. The test subject is attracted to the target and looks at it in preference to the backdrop. When the grating is too fine to be resolved, the target becomes

indistinguishable, blending into the backdrop, and preferential looking is no longer activated. The detection of a grating relative to background underpins Lea Paddles (LP, Lea-Test Ltd and Good-Lite, Illinois, USA, demonstrated in Figure 1A), Keeler Acuity Cards for Infants (KACI, Keeler, Windsor, UK, demonstrated in Figure 1B), Teller Acuity Cards (TAC, Vistech Consultants Inc. Ohio, USA, demonstrated in 1C), and the computer tablet based digital acuity test, Peekaboo Vision (PV, Scottish Health Innovations Ltd, Clydebank, UK, demonstrated in Figure 1F). The iterative design and evaluation of PV is under evaluation as part of the present Thesis.

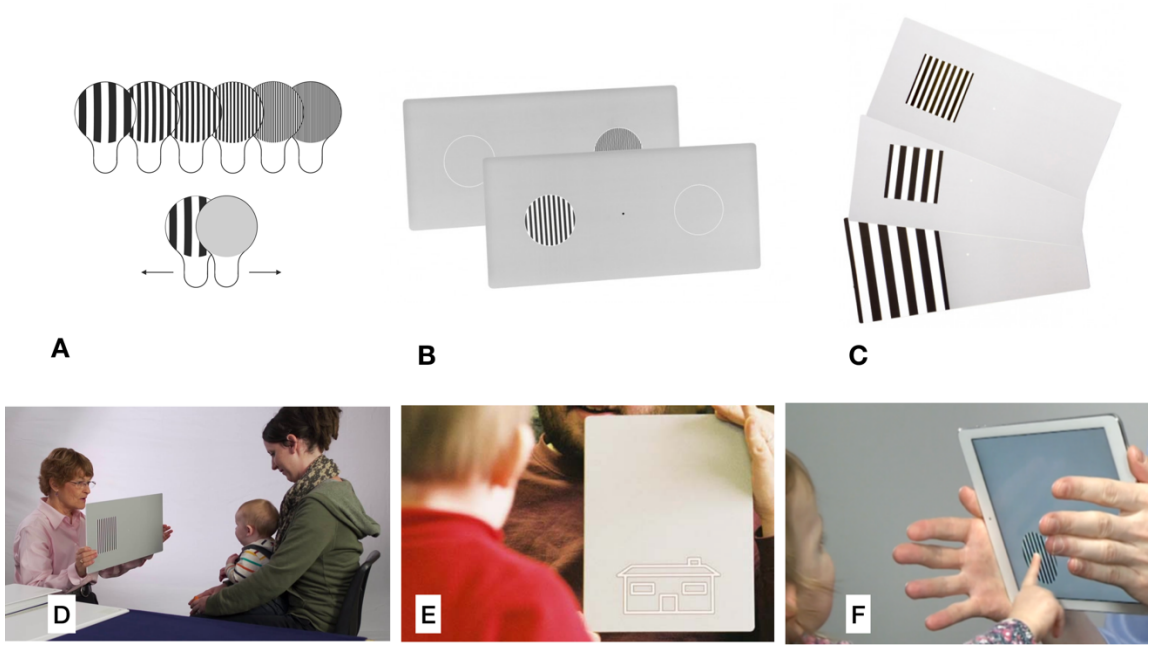


Figure 1. Traditional Card Based Platforms

1A. Lea Paddles [19]; 1B. Keeler Acuity Cards for Infants [20]; 1C. Teller Acuity Cards [21]; 1D. Teller Acuity Cards (demonstrates typical use) [22]; 1E. Cardiff Acuity Cards [23]; 1F. Peekaboo Vision on iOS

Fundamental to the paradigm, the luminance of the grey background is required to precisely match the average luminance of the high contrast gratings. Cardiff Acuity Cards (CAC, Good-Lite, Illinois, USA, demonstrated in Figure 1E) employ the *vanishing optotype* paradigm, first described by Howland et al [24], where an outline of a

recognisable target (in the case of the format adopted by present incarnation of CAC) a car, duck, boat, apple or house is generated by composite lines where the central white line is bordered by a dark outline half the thickness of the white line (demonstrated in Figure 2), effectively generating pseudo high-pass design that can be used to infer a mean angle of resolution, which is typically expressed as a Snellen equivalent value. Unique to CAC, in addition to *detection* functionality, the test serves as *discrimination* test also, and is not solely a FCPL test.

While amblyopia in infancy is typically evaluated within the context high contrast spatial frequency acuity norms in terms of forced choice preferential looking testing, vernier discrimination has been reported to precede that of grating acuity [25]. Indeed, while grating resolution is a singular aspect underpinning an acuity score, the threshold perceived to have been reached by a child during a typical acuity test is a behavioural response to the summation of multiple visual functions. The impact of amblyopia upon contrast sensitivity [26], stereopsis [27], motion processing [28] and colour vision [29] remains largely an academic pursuit, beyond the clinician's *real life* diagnostic armamentarium. Differing aetiologies are likely to impart differing severities across distinct neurologic physiologies. This is exemplified by the relative lack of recovery of stereopsis (despite improved acuity scores) in *strabismic* amblyopes versus *anisometropic* amblyopes, whose stereopsis appears relatively improved in comparison.

While strict national and international regulatory standards apply to *adult* acuity tests (Reviewed in related studies: Livingstone et al, 2016), to our knowledge, despite the time-critical phase of visual development under test, there are no regulatory standards applicable to grating/vanishing optotype style infant vision tests. The impact of a deviation in terms of target luminance upon performance of the test is not presently clear, and to our knowledge, has not been formally evaluated within the literature.

Furthermore, there are several paradigms for FCPL acuity, all of which have differences in the detection task and staircasing method.

Staircasing refers to the protocol adopted for progressing through an acuity test, whereby presenting an easier visual target reflects a step *up* the staircase, whereas presenting a harder visual target (finer grating in this case), reflects a step *down* the staircase. A *reversal* indicates a change in direction *up* or *down.*

Although the mean angle of resolution of the lines/gratings dictates the acuity *score* (typically expressed as logMAR), acuity results appears to vary widely between FCPL tests, and it is generally accepted that grating acuity results cannot be accepted as interchangeable regards tests (TAC/CAC/KAC/LP). TAC are reported to have advantages over CAC for detection of amblyogenic conditions in a young population[30]. Comparing with distance letter optotype discrimination acuity, TAC was reported to underestimate the presence of amblyopia, with a negative predictive value for detection of visual deficit (20/70 or poorer) of 50%, and for legal blindness (20/200), 71%[31].

When CAC were compared with LP in an Indian cohort [32] of 230 children (aged 2-36 months), the difference in acuity level between the two test modalities were found to be highly statistically significant ($p<0.0001$, students t test), though correlation was felt to be strong (Pearson's Correlation Coefficient 0.98 binocularly, 0.62 monocularly). Nevertheless, the difference in results between tests was felt to preclude interchangeable use.

Shah et al [33] evaluated digital vanishing optotype *letter* targets in three normal vision adult volunteers, and found the vanishing optotype paradigm presented a promising alternative to traditional sloan style high contrast, ETDRS style letter optotypes, with a pronounced decrease in measurement variability compared with the traditional standard.

Testing was performed with a gamma-corrected high-resolution CRT (Cathode Ray Tube) monitor (background luminance of 53.9 cd m2). A bilinear interpolation computer graphics technique was used to display achieve sub-pixel resolution of the vanishing optotypes.

When approaching the limits of visual perception, converging multi-modal inputs to the eye are subject to complex retinal, brainstem and cerebral visual processing, culminating in the reflex detection of exquisitely subtle stimuli. A deviation of the grey backdrop from the target average luminance of the test gratings risks a *contrast difference* between grating and backdrop, potentially driving a looking response, leading to a result being mistakenly interpreted as high contrast acuity perception when it actually reflects a (potentially gross) contrast sensitivity function.

Furthermore, light reflected from printed cards and illuminated light from a screen comprise a range of spectral wavelengths. While nominally, the target and backdrop features within a grating test are attributed black, white or grey values, potential exists for inadvertent weighting towards a particular colour dominant, which may vary for each feature, creating another potential confounding factor when interpreting a preferential looking result. In addition, for card-based platforms particularly, such reflected light properties are implicitly related to the spectral characteristics of the ambient illumination, which can vary significantly within test settings, from fluorescent clinic lighting, to LED (Light Emitting Diode) or incandescent bulbs, to daylight, or frequently combination lighting.

1.1.3 Infant Vision Testing Standards and Evolving Digital Platforms

The widespread use of mobile technology is generating innovative ways to improve health. Globally, it is estimated that 500 million mobile device users will download healthcare applications ('apps') by within 5 years [34]. This brings an opportunity to enhance detection of compromised vision.

Given today's knowledge, technology and treatment, and that an estimated 80% of global blindness is preventable or curable [35], mobile technology-based approaches represents an obvious and credible means of radically changing how we detect visual disability [36,37]. It is therefore unsurprising that mobile technology, in particular the iPad tablet (Apple Inc., CA, USA), has been evaluated as an alternative to traditional chart-based methods for measuring vision [38–42]. Aslam et al [38] investigated the fundamental physical characteristics of the iPad 3 display in relation to vision testing. Their results suggested that while the tablet screen was unable to exactly match the low levels of contrast in the Pelli-Robson contrast sensitivity chart, the device has the potential to screen for contrast sensitivity defects across a broad range. Investigation of the impact of viewing angle similarly found that despite relatively high absolute changes in luminance, contrast remained highly stable, suggesting minimal significant impact on clinical testing.

However, only a minority of apps presently available have been subject to robust evaluation, and the number of apps is rising: in 2012, 32 iPhone (Apple Inc., CA, USA) apps purported to assess visual function [42], while currently (23/06/2019), the search

phrase “vision test” in Apple’s online app store (www.store.apple.com/uk) generates 202 results.

As with more traditional medical devices, apps used in a clinical context, and the platform or device on which they run [43], require regulation because of potential risks to the public. Updated guidance regarding which applications and target platforms are appropriate for clinical use is needed, and regulation, accreditation and ‘kitemarking’ of various healthcare technologies including apps is planned in the UK [44].
The present thesis presents research around the physical properties of traditional and ubiquitous digital platforms.

Given that the present reference standard for infant vision remains unchanged over the past 35 years [14,15,45]: a printed surface, we seek to evaluate the place of emerging tablet screen technologies in the context of acuity measuring medical devices.

Guided by these observations, a prototype infant acuity application was developed, refined and tested in adult, then paediatric groups across low/middle and high resource settings. These studies aim to robustly establish utility of tablet computer-based infant testing as a (potentially global) clinical resource towards a validated tool to help detect poor vision as early as possible in childhood, with a view to connecting children with visual problems to validated treatments that can limit life-long visual deficits.

Although amblyopia is the most common reason for poor vision in childhood, and the most obvious target for a visual screening technology, the objective of the present development and clinical evaluation relates to evaluation of a new digital format of

paediatric vision, evaluating whether such a platform can be used as an alternative to present printed standards, irrespective of the underlying pathology.

Within this work we evaluate the physical properties of digital and card-based platforms, and compare results from traditional and digital index tests (prototypes and completed builds) across paediatric groups.

The hypothesis under test is that the computer tablet platform offers intrinsic advantages relating to photometric properties required for mounting an acuity test, when compared to printed card-based standards. We hypothesise that, compared to the printed standards, the digital platform can more precisely attain a target isoluminant backdrop: a property central to conventional forced-choice preferential looking paradigms.

Furthermore, together with the ability to offer rapid transitions, animation/sound rewards, and exert an automated staircasing paradigm, an appropriately designed test can provide advantages in terms of time-to-test and test-retest repeatability.

1.2 Thesis structure

The present research is presented in distinct 5 consecutive parts, with the core studies divided into chapters 2 to 5, focussing first on the intrinsic physical properties of the devices, and performance relative to intended use, and subsequently *real life* testing with a focus on acceptability, accuracy and performance.
The order of presented chapters has been chosen to provide a logical narrative but does not reflect the chronological order in which the studies were performed.

1.2.1 Photometric Compliance and High Contrast Acuity Testing: Retro-Illuminated Charts and Computer Tablets

The most prescriptive, and tightly standardised for of vision testing in clinical practice is the high contrast acuity chart, specifically, the ETDRS chart. Hence, this is the target we use to measure against with alternative digital platforms, which we speculated would have the potential to extend the traditional armamentarium of robust vision testing tools, with the potential to achieve the high standards applied to adult high contrast vision, and extend them to infant testing.

Chapter 2 evaluates the physical properties of tablet devices regards fitness for purpose regards visual acuity measurement.

Detailing National and International standards for high contrast acuity chart design, luminance, contrast and luminance uniformity for both the index platforms (tablet devices) and the traditional charts (ETDRS), we evaluated under different ambient lighting conditions.

1.2.2 Shades of Grey: Photometric Compliance of Digital Platforms and Traditional Infant Acuity Cards

Chapter 3 develops the theme of conformity to standards further, applying these adult retro-illuminated chart standards to the leading infant *card-based* testing platforms, as well as a wider range of digital displays.

Achieving an average grey luminance that matches the mean average luminance of the composite black/white foreground elements is a cardinal aspect of the forced-choice preferential looking test paradigm, yet to our knowledge, no standards or recommendations exist to ensure this is achieved. The photometric properties of the target iso-luminance of the background is investigated across traditional and digital

platforms. The (non-Lambertian) traditional platforms are additionally evaluated under more extensive illimitation conditions: hot/cold studio light, fluorescent light and ambient light.

1.2.3 Digital Grating Acuity Testing in an Adult Cohort: A Proof of Concept

With all platforms undergoing a systematic evaluation of their photometric properties in the context of existing standards, we then aim to establish how the digital platform measures against reference standards in terms of clinical testing. Chapter 4 details early the design of a prototype digital infant acuity test, Peekaboo Vision Prototype Build 1 (PVb1) leveraging high- spatial frequency gratings. In a cohort of adult volunteers with normal vision, comparisons are made with the leading grating test in the UK, the Keeler Acuity Cards for Infants (KACI, Keeler, Windsor, UK). In this proof of concept phase, performance at low levels of vision was simulated using a range of spherical plus lenses. This allowed a large number of tests across a relatively small group, providing enough power to undertake Bland Altman analysis, as well evaluation of test-retest repeatability.

1.2.4 Digital Grating Acuity Testing in Paediatric Cohorts

Once potential utility had been evaluated in an adult cohort, the next studies aimed to evaluate performance in a paediatric cohort, comparing with the same reference standard, KACI. Chapter 5 combines 2 studies spanning 2 builds of the app across 2 contrasting settings, evaluated consecutively in Studies 1 and 2.

1.2.4.1 Study 1

Peekaboo Vision Prototype Build Version 1 (PVb1) was first evaluated in rural African setting, under comparison with KACI in a young cohort, with analysis of time-to-test,

accuracy and repeatability. Given the prototype nature of the application build, Study 1 was also treated as a bug-finding phase to inform a subsequent build in advance of a more formal evaluation of the technology. Indices of compliance were also evaluated for both platforms and compared. The methodology for the study was also under test, making Study 1 pilot in nature in advance of a more formal evaluation, Study 2.

1.2.4.2 Study 2

Informed by Study 1, changes were made to the app format as well as the testing methodology. Peekaboo Vision, formal build for iOS (PVi), was evaluated in a UK setting in a similar study in most respects (6-60 months age range).

Discussion of the results for Studies 1 and 2 are combined in Chapter 5 to provide a more meaningful and less repetitive narrative.

The final section, Chapter 6, summarises the main findings from the presented research, and outlines areas for future research and development.

Chapter 2: Photometric Compliance of Tablet Screens and Retro-illuminated Acuity Charts as Visual Acuity Measurement Devices

2.1 Introduction

High contrast visual acuity assessment remains the key, measurable outcome for defining abnormal vision and mapping changes in visual function, and has been widely targeted for apps on mobile devices [36,37,41,46,47]. Both optotype contrast and test luminance affect acuity measurements [48,49], and therefore devices purporting to measure acuity for clinical purposes should be standardised as specified for chart tests.

Chart luminance should be ≥ 80 cd/m^2 [50] or ≥ 120 cd/m^2 [51], depending on the standard used. The ETDRS specification of 160 cd/m^2 is based on the ICO (International Council of Ophthalmology) recommendations [50]. An evidence-based recommendation of 80–320 cd/m^2 [48] was adopted for a non-clinical international standard [52]. Contrast specification also differs by standard: the ICO standard specifies that black optotypes should be ≤ 15% of the luminance of the white surrounding field [50], while BS 4274-1:2003 specifies that luminance (Weber) contrast [$(L_{bkg} – L_{letter}) / L_{bkg}$] should be ≥ 90%. Luminance uniformity is specified only in BS 4274-1:2003, which requires that any variation across the chart should be ≤ 20%.

Previous work has demonstrated suitable physical screen characteristics of three tablets from different manufacturers for contrast vision testing, providing individual device gamma functions are known [38,40]. The authors also noted that luminance uniformity varied within and between devices, but with little detriment to contrast uniformity. This current study aimed to develop this work for high contrast acuity testing. It is likely that gamma functions matter less for high contrast acuity, and given the need for simple but sensitive point-of-care test procedures, more complex

calibration procedures are not desirable. We investigated seven tablets (iPads, Apple Inc., CA, USA), measuring luminance, contrast and luminance uniformity and the effect of adjusting brightness settings. The tablets selected allowed comparison within the same device generation as well as across three generations of the currently available iPad range. We also assessed luminance, contrast and luminance uniformity of three gold standard ETDRS visual acuity charts and compared these with the tablet findings. The effect of room lighting (on or off) was investigated for a subset of tablets and all three charts.

2.2 Methodology

Seven tablets were evaluated (one iPad 3, three iPad 4s, three iPad Air 2s, Apple inc, California, USA) along with three ETDRS charts (Precision Vision, IL, USA) mounted in separate illumination cabinets (Sussex Vision, Rustington, UK) (See Figure Figure 2. The three ETDRS charts). The charts and illumination cabinets varied in age, but were all in clinical use. The illumination cabinets enclosed 2 horizontal bulbs, illuminating interchangeable letter optotype test screens.

Luminance measurements were performed using a Minolta LS-100 luminance meter (Konica Minolta Sensing, Europe B.V.) with a 1° aperture and a No. 110 close-up lens with a current calibration traceable to the German national standard. The procedure was similar to those recommended by national and international protocols [53,54].

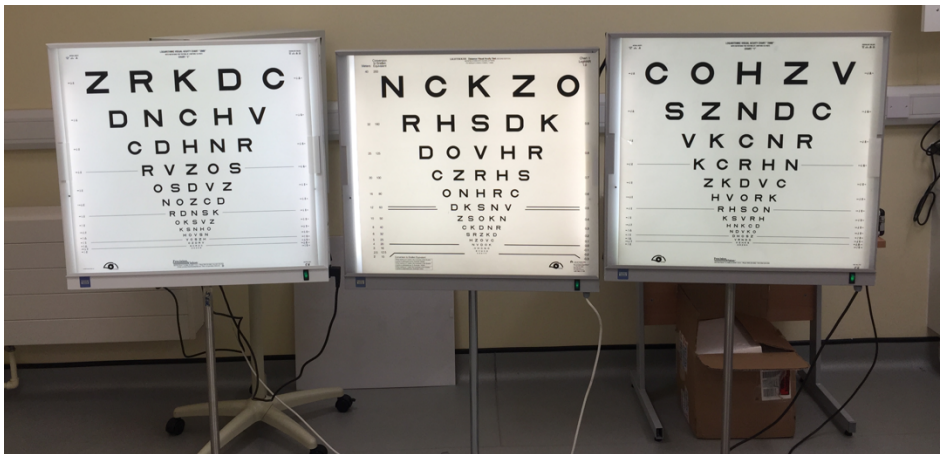


Figure 2. The three ETDRS charts

At the time of writing, technical information on typical maximum brightness is available for the iPad Air 2 (2014 edition), quoted at 400 cd/m2 max brightness (typical), as published online by Apple[55]. For the iPad 3 and iPad 4. generation under devices test, the display properties are unavailable on the Apple website, but have been

quoted by third parties at similar levels for the iPad 3 (394 cd/m2) and iPad 4 (407 cd/m2)[56].

Each tablet was placed on a horizontal surface with the luminance meter held in a tripod and positioned perpendicular to the screen. The meter was brought into focus on the position of interest and the distance between the screen and lens measured at 55 +/- 5 mm. For optimum screen stability, the tablet devices were switched on at least 15 minutes prior to data acquisition [38]. Auto-brightness was deactivated. Ambient light levels were measured with the Iso-tech ILM-01 light meter (RS Components, UK). Room lights were switched off to avoid reflections and stray light [52], with ambient light levels measuring 0.52 lx. To emulate real-life test conditions, three of the tablets were also measured with the room lights on (the 3 tablets were from the iPad Air 2 range), where ambient light levels measured 762 lx.

A two-frame reversing black and white checkerboard test screen was created in Adobe Photoshop CC v14 (Adobe Systems Inc., CA, USA) with 186×186 pixel black (RGB 0 0 0) and white (RGB 255 255 255) squares. This allowed both maximum and minimum luminance measurements at the same screen location. Measurements were made at nine cardinal points (Figure 3), moving the screen with respect to the photometer between each point to maintain the perpendicular aspect [53].

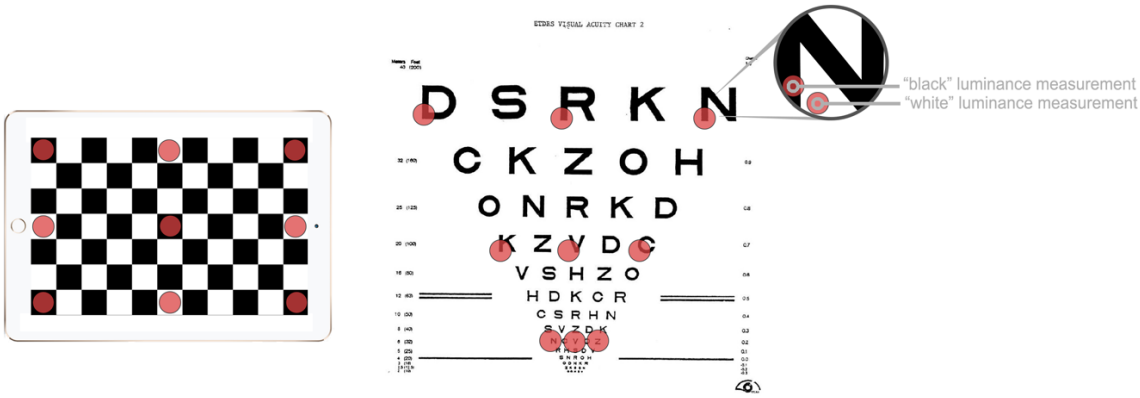


Figure 3. Illustration of cardinal points (red dots) where measurements were made for the tablets (left) and for the EDTRS charts (right)

Five levels of screen brightness were used based on linear position of the brightness setting slider bar, measured with a ruler. Minimum ("0%") and maximum ("100%") used the extreme positions, and three evenly spaced intermediate positions were used, hereafter labelled "25%", "50%", and "75%". A manual brightness setting method was chosen over software-based methods, so if found necessary, brightness settings could be easily changed by health care professionals without dependence on bespoke software.

ETDRS charts were switched on for at least one hour prior to taking measurements. The luminance meter was hand-held 60 +/- 5mm perpendicular to the chart and focused on the lower left aspect of the black target letter, and on its adjacent white background. Room lights were switched on throughout to emulate routine clinical use: there is no guidance regarding calibration with or without additional room lighting. We therefore repeated the measurements on all charts with the room lights off.

For the seven tablets, each at five different brightness settings, and for the three EDTRS charts, overall luminance was calculated as the mean of the nine measurements

across the screen or chart. Contrast was calculated for each of the nine cardinal points of the tablets by comparing luminance when black (L_{black}) with luminance when white (L_{white}), and for each optotype of the charts by comparing luminance of letters (also L_{black}) with adjacent white background (also L_{white}). Michelson contrast [$(L_{white}-L_{black}) / (L_{white}+L_{black})$], a simple contrast ratio (L_{black}/L_{white}), and Weber contrast [$(L_{white}-L_{black}) / L_{white}$] were calculated. The simple contrast ratio was calculated to determine compliance with the ICO standard which requires a ratio of ≤ 15% [50], and Weber contrast was calculated to determine compliance with BS 4274-1:2003, which requires a Weber contrast ≥ 90% [51]. Luminance uniformity was calculated as the ratio of the lowest of the nine white measurements to the highest (L_{min}/L_{max}) and used to determine compliance with BS 4274, requiring a ratio ≥ 80% [51].

Pilot testing established that variation of luminance within the black/white squares and within the chart letters/proximal background were minimal. Similarly, varying the focal distance of the luminance meter within ± 5 mm had negligible impact on luminance values.

2.3 Results

2.3.1 Luminance and contrast

For all seven tablets, mean luminance of the display decreased as the brightness setting of the tablet was reduced, as expected. For white squares, luminance dropped from around 300 cd/m^2 at the maximum setting to around 3.5 cd/m^2 at the minimum setting, while black square luminance dropped from around 2–3 to 0.02 cd/m^2 (

Table 1, Figure 4).

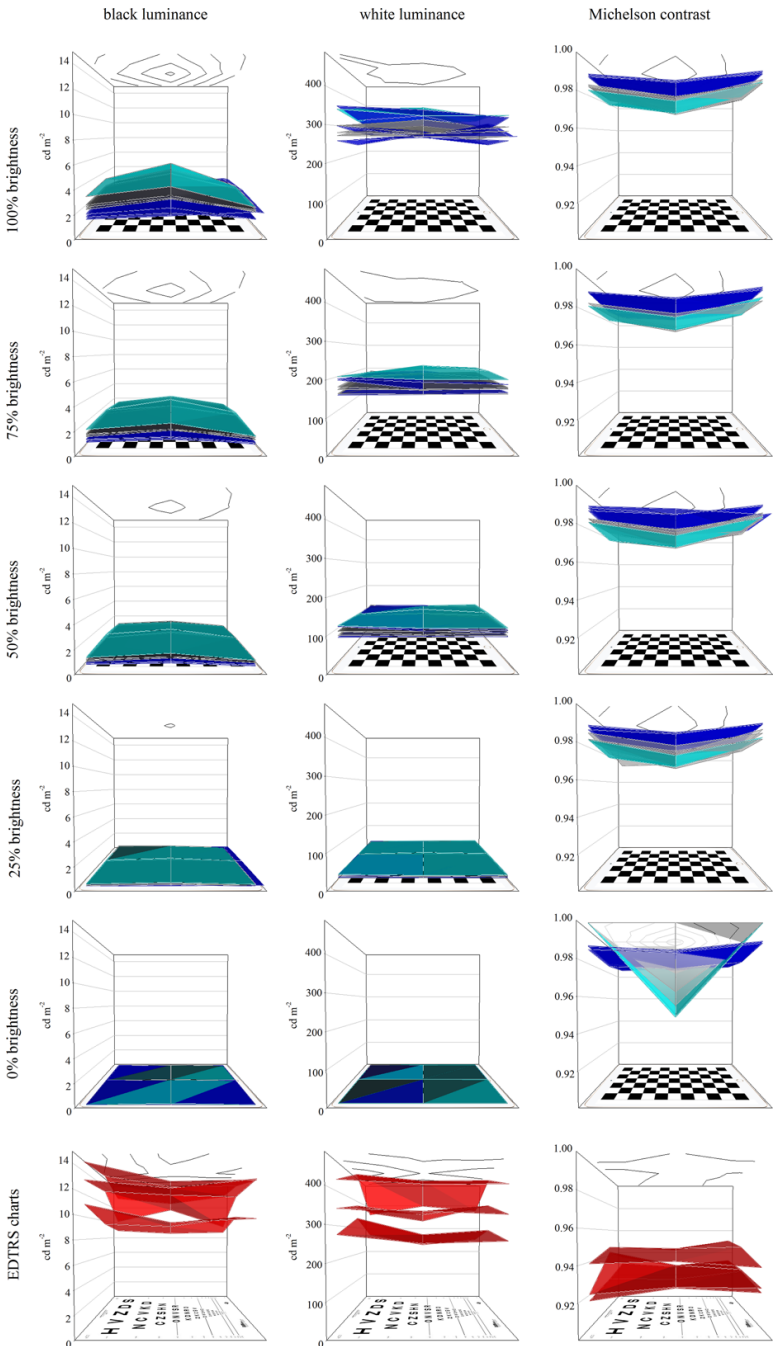


Figure 4. Luminance Surface Plots

Luminance of black (left) and white (centre) squares, and of Michelson contrast (right) for the seven tablets at nominal maximum ("100%", top) to minimum ("0%", bottom) brightness settings. Coloured surfaces in each plot also demonstrate luminance and contrast uniformity, by joining the nine measured values. Ceiling contour plots represent the mean uniformity profile of all seven tablets. Cyan surface: iPad 3 tablet. Grey surfaces: iPad 4 tablets. Blue surfaces: iPad Air 2 tablets. Data for the three EDTRS charts based on luminance of black optotypes and adjacent white backgrounds are shown in the lowermost plots (red surfaces).

Mean Michelson contrast remained almost unchanged with brightness setting (97.9–99.6%). At minimum ("0%") brightness, the luminance of the black squares measured zero in one case, created a spurious 100% contrast (Table 2, Figure 4, bottom right panel). Newer generations of tablet had lower black square luminance ("blacker blacks") than older generations: iPad Air 2 black squares were typically 0.2 log units less bright than iPad 4 black squares. White square luminance changed much less: typically iPad Air 2 white squares were 0.02 log units less bright than iPad 4 white squares. Consequently, contrast was slightly higher for the iPad Air 2 tablets: ignoring values for the minimum ("0%") brightness, iPad Air 2 Michelson contrast was 98.7–98.8% for all devices and four brightness settings, while iPad 4 Michelson contrast was 97.9–99.2% (Table 2)

The mean white luminance of each EDTRS chart varied widely and black letter luminance always exceeded the luminance of tablet black squares, even at maximum tablet brightness setting. Consequently, mean Weber contrast for each chart (96.2–97.1%) was lower than for any tablet/brightness setting combination (Table 3).

2.3.2 Compliance with standards

All tablets met the overall luminance requirement of the ICO (≥ 80 cd/m^2) providing the brightness setting was 50% (half-way) or higher. A nominal brightness setting of 75% or higher was required to meet the BS 4274 requirement (≥ 120 cd/m^2), although at the 50% setting, mean luminance was only around 0.05 log units below the required level (

Table 1). All EDTRS charts met the requirements of both standards (Table 3).

Contrast of the tablets across all brightness settings was $\leq 1.04\%$ in terms of the contrast ratio (L_{black}/L_{white}) used by the ICO standard, exceeding its requirement to be $\leq$

15%. Weber contrast was ≥ 99% for tablets across all brightness settings, exceeding the minimum requirements (≥ 90%) of BS 4274-1:2003. EDTRS charts had slightly poorer contrast, with contrast ratios (L_{black}/L_{white}) of 3.8%, 3.5% and 2.9% and Weber contrasts of 96.2–97.1%, but still exceeding contrast requirements of both standards.

2.3.3 Luminance uniformity

Luminance uniformity was generally high for the tablets (room lights off), and was not greatly affected by the brightness setting. The most uniform device (iPad 4 #1) showed less than 10% change in luminance across its surface, whilst the least uniform device, iPad 4 #2, showed around 23% change in luminance across its surface (Table 4, Figure 3). The EDTRS charts (room lights on) all showed greater luminance variation than even the most variable of the tablets: changes in luminance from the brightest to the dimmest part of the chart background measured 15%, 30% and 25% for ETDRS charts 1, 2 and 3 respectively (Table 3, Figure 3).

2.3.4 Comparison with standards

Luminance uniformity of the tablets was very close to or exceeded the requirements of BS 4274 ($L_{min}/L_{max} > 80\%$). Two of the seven devices did not meet the requirements: one (iPad 4 #2) had uniformity of 76–77% across the five brightness settings, and another (iPad Air 2 #1) had uniformity of 77–81% across the five brightness settings (Table 4).

nominal brightness setting	100%		75%		50%		25%		0%	
	black	white	black	white	black	white	black	white	black	white
iPad 3	3.3 (1)	331 (17)	2.0 (0.6)	202 (11)	1.2 (0.3)	**118 (7)**	0.37 (0.1)	**36 (2)**	0.03 (0.0)	**3.9 (0.2)**
iPad 4 #1	2.8 (0.7)	287 (10)	1.7 (0.4)	178 (6)	1.1 (0.3)	**111 (3)**	0.34 (0.1)	**33 (1)**	0.02 (0)	**3.6 (0.1)**
iPad 4 #2	2.6 (0.7)	278 (21)	1.6 (0.4)	165 (13)	0.9 (0.2)	**100 (8)**	0.28 (0.1)	**28 (2)**	0.01 (0)	**3.4 (0.3)**
iPad 4 #3	2.8 (0.8)	300 (13)	1.7 (0.5)	182 (8)	1.0 (0.3)	**106 (5)**	0.30 (0.1)	**33 (2)**	0.00 (0)	**3.6 (0.2)**
iPad Air 2 #1	2.0 (0.5)	315 (26)	1.2 (0.3)	181 (14)	0.7 (0.2)	**108 (8)**	0.21 (0)	**32 (3)**	0.02 (0)	**3.5 (0.2)**
iPad Air 2 #2	1.8 (0.5)	270 (20)	1.2 (0.4)	179 (13)	0.7 (0.2)	**110 (8)**	0.21 (0.1)	**32 (2)**	0.02 (0)	**3.3 (0.2)**
iPad Air 2 #3	1.9 (0.4)	281 (16)	1.1 (0.2)	166 (9)	0.6 (0.2)	**96 (5)**	0.20 (0.1)	**30 (2)**	0.02 (0)	**2.8 (0.1)**

Table 1. Luminance values (cd/m^2) for seven tablets at five different nominal brightness settings.

Values are the mean of the nine values measured across the screen for black and for white squares. Standard Deviation in brackets. Measurements were made with room lights off. Device/setting combinations for white squares which do not meet the BS 4274 requirement (≥ 120 cd/m^2) are in bold, and those which do not meet the ICO requirement (≥ 80 cd/m^2) are underlined.

nominal brightness setting	100%			75%			50%			25%			0%		
	MC	L_{black} / L_{white} %	WC	MC	L_{black} / L_{white} %	WC	MC	L_{black} / L_{white} %	WC	MC	L_{black} / L_{white} %	WC	MC	L_{black} / L_{white} %	WC
iPad 3	98.0	1.00	99.0	99.0	1.00	99.0	97.9	1.04	99.0	98.0	1.02	99.0	98.7	0.65	99.4
iPad 4 #1	98.0	0.98	99.0	99.0	0.98	99.0	97.9	1.00	99.0	98.0	1.04	99.0	98.7	0.60	99.4
iPad 4 #2	98.1	0.94	99.1	99.1	0.94	99.1	98.2	0.92	99.1	98.1	0.97	99.0	99.6	0.20	99.8
iPad 4 #3	98.2	0.92	99.1	99.2	0.92	99.1	98.2	0.90	99.1	98.2	0.92	99.1	100	0.00	100
iPad Air 2 #1	98.8	0.62	99.4	98.7	0.65	99.4	98.7	0.64	99.4	98.7	0.64	99.4	98.6	0.70	99.3
iPad Air 2 #2	98.7	0.67	99.3	98.7	0.66	99.3	98.7	0.64	99.4	98.7	0.63	99.4	98.6	0.70	99.3
iPad Air 2 #3	98.7	0.67	99.3	98.7	0.66	99.3	98.7	0.66	99.3	98.7	0.68	99.3	98.5	0.74	99.3

Table 2. Contrast values (%) for seven tablets at five different nominal brightness settings

Measurements were made with room lights off. Values are the mean of the nine calculated contrasts from values measured across the screen for black and for white squares. MC: Michelson contrast. WC: Weber contrast. All tablets met the requirements of the ICO standard ($L_{black}/L_{white} \leq 15\%$) and of BS 4274 (Weber contrast $\geq 90\%$).

	black optotype luminance (cd/m^2)	adjacent white luminance (cd/m^2)	contrast L_{black}/L_{white}	Weber contrast	luminance uniformity L_{min} / L_{max} (%)
chart #1	13	354	3.8	96.2	85
chart #2	9.5	270	3.5	96.5	**70**
chart #3	12	418	2.9	97.1	**75**

Table 3. Luminance (cd/m^2), contrast (%) and luminance uniformity of three EDTRS charts.

Measurements were made with room lights on. Values represent the mean of the nine values measured across the chart for black optotypes and for adjacent white areas. All charts met BS 4274 (≥ 120 cd/m^2) and ICO requirements (≥ 80 cd/m^2) for luminance and for contrast (BS 4274, Weber contrast ≥ 90 %; ICO, L_{black}/L_{white} ≤ 15 %). Chart parameters which do not meet the BS 4274 luminance uniformity requirements (>80 %) are in bold.

nominal brightness setting	100%	75%	50%	25%	0%
iPad 3	84	83	80	83	82
iPad 4 #1	91	91	91	89	90
iPad 4 #2	**77**	**77**	**76**	**77**	**77**
iPad 4 #3	85	86	85	85	82
iPad Air 2 #1	**77**	**78**	**78**	**77**	81
iPad Air 2 #2	80	81	80	80	80
iPad Air 2 #3	85	85	86	85	88

Table 4. Luminance uniformity (white squares, L_{min}/L_{max}, %) of seven tablets at five different nominal brightness settings. Measurements were made with room lights off.

Device/setting combinations which do not meet the BS 4274 requirements (> 80%) are in bold.

2.3.5 Effect of room lighting

Three tablets were re-measured, but this time with the room lights on to emulate a typical visual acuity testing situation. Nominal brightness setting of 75% was used. Switching on the room lights increased the apparent average white luminance of the tablets by about 5% (0.02 log units), but increased the average black luminance by over 200% (over 0.5 log units). Michelson contrast and Weber contrast fell by 3% and by 1–2% respectively, while L_{black}/L_{white} contrast ratio increased by 1–2%. Luminance uniformity improved a little for all tablets, bringing all into BS 4274 specification (>80%) (Table 5).

All three charts were re-measured with the room lights off. This effectively measures the intrinsic luminance of the devices, decreasing the apparent average white luminance of the EDTRS charts by 32% (0.17 log units), and decreasing the average optotype luminance by 59% (0.38 log units). All charts continued to meet both the ICO (≥ 80 cd/m^2) and BS 4274 (≥ 120 cd/m^2) requirements. Weber contrast was slightly higher with the room lights off. With room lights *on*, luminance uniformity improved for two charts, bringing one chart into BS 4274 specification (>80%), and was unchanged for the third chart (Table 6).

	black square luminance (cd/m^2)	white square luminance (cd/m^2)	Michelson contrast	contrast L_{black}/L_{white} (%)	Weber contrast (%)	luminance uniformity L_{min}/L_{max} (%)
iPad Air 2 #1	1.2→4.3	181→191	98.7→95.5	0.65→2.3	99.4→97.7	**78**→82
iPad Air 2 #2	1.2→4.3	179→190	98.7→95.5	0.66→2.3	99.3→97.7	81→81
iPad Air 2 #3	1.1→3.8	166→174	98.7→95.7	0.66→2.2	99.3→97.8	85→89

Table 5. Pairs of data (room lights **off**→room lights **on**) for three tablets.

For luminance and contrast, values represent the mean of the nine values measured across the screen for black and for white squares at the same location, with the nominal brightness level set at "75%". Screen parameters which do not meet the BS 4274 requirements are in bold.

	optotype luminance (cd/m^2)	white luminance (cd/m^2)	contrast L_{black}/L_{white} (%)	Weber contrast (%)	luminance uniformity L_{min}/L_{max} (%)
chart #1	5.2→13	244→354	2.1→3.8	97.9→96.2	**76**→85
chart #2	3.1→9.5	174→270	1.8→3.5	98.2→96.5	**66→70**
chart #3	6.6 →12	301→418	2.2→2.9	97.8→97.1	**75→75**

Table 6. Pairs of data (room lights **off**→room lights **on**) for EDTRS charts.

For luminance and contrast, values represent the mean of the nine values measured across the chart for black letters and for adjacent white areas. Chart parameters which do not meet the BS 4274 requirements are in bold.

2.3.6 Summary

Two standards describe the photometric qualities required of visual acuity test charts [50,51]. We have shown that seven tablets, measured in a dark room, exceeded the standards' requirements for mean luminance providing the nominal brightness setting is above 50%, and exceeded the standards' requirements for contrast regardless of brightness setting. Two of the seven tablets fell marginally short of the required luminance uniformity threshold. Re-assessing three tablets at a nominal 75% brightness setting and with room lights on made little difference to mean luminance or to contrast, but all three tablets then exceeded the luminance uniformity threshold, where one had previously failed. We have also shown that three typical, clinical standard ETDRS charts in a well-lit room exceeded the standards' requirements for mean luminance and for contrast, but two charts fell short of the required luminance uniformity threshold. With room lights off, mean luminance and contrast remained adequate, but all three charts then failed to meet the luminance uniformity requirement. Tablets showed much less inter-device variability, higher contrast, and (under room lighting) better luminance uniformity than charts, providing they were operated at suitably high brightness.

2.4 Discussion

The iPad tablets tested here were more compliant with British and international photometric standards for vision testing than the retro-illuminated ETDRS charts currently used in a tertiary referral dedicated ophthalmic unit. Standards for luminance and for contrast were met by both tablets and charts, but charts generally had lower contrast and greater variability in both luminance and contrast. The mean luminance fluctuations measured here (for example, tablets at 75% brightness varied from 165–202 cd/m^2, and charts ranged from 270–418 cd/m^2) are very unlikely to affect clinical measurement: even doubling chart luminance in this range improves acuity by just 1 letter optotype in a line [48,57]. Weber contrast varied between tablets (at 75% brightness, for example) from 99.0–99.4%, and varied between charts from 96.2–97.1%.

Luminance uniformity was 77–91% for the tablets (75% brightness, room lights off), in good agreement with findings elsewhere for an iPad 3 screen of variation from -5 to -23% relative to screen centre [38]. Luminance uniformity was slightly poorer and more variable for the charts (70–85%, room lights on). The adverse effect of uneven luminance would be to create zones where the contrast of the optotype relative to its background differs substantially from other areas of the test surface, introducing greater error in clinical thresholds, or even elevating thresholds. However, for both tablets and charts, the cause of the luminance non-uniformities (uneven-ness of either intrinsic screen brightness of extrinsic lighting) affects both black and white areas, and contrast is thus relatively unaffected (see Figure 4, right hand diagrams), as found elsewhere [47]. Indeed, given the plateau observed in the relationship between luminance and acuity within the range of chart luminance, the stringent criterion referring to uniformity in BS 4274-1:2003 is difficult to explain from a clinical perspective for high contrast acuity. However, this standard may be relevant for contrast sensitivity testing, where the additive effects of luminance and contrast are important, and large discrepancies are

reported in the test results between testing modalities, particularly under varying lighting conditions [58].

Study limitations

We made no effort to control for the effects of battery power. Other studies have shown screen luminance stability of the first generation iPad, maintaining 275 cd/m^2 from 100% to 5% battery charge [41]. For the iPad 3 on battery power, only a small luminance loss was detected immediately after switching the device on, which was more pronounced if the device was switched off for longer periods [38].

We did not evaluate variability in luminance with viewing angle, instead measuring luminance perpendicular to the tablet or chart surface [53]. Viewing angle is of particular relevance to the tablet platform, given angular effects are likely to be more pronounced when testing is at a closer range. This parameter has been investigated by Parry et al, who tilted photometric devices to mimic the changing viewing angle of the human eye. Testing was performed on the iPad 3 (Apple inc) [38] and also a Google Nexus 10 (Google Inc.) and Galaxy Tab 2 10.1 (Samsung Electronics) [40]. The mean contrast changes over the peripheral areas of the screen varied between devices, but in terms of contrast, the impact was minimal, at most around 1%. The authors asserted that this would be unlikely to be clinically perceptible, as it is less than one just-noticeable difference (JND).

It was not the aim of the current study to undertake any acuity measurements, but it is probable that tablet screen reflections do present an obstacle not present with ETDRS charts. A masked diagnostic study of visual acuity using a first generation iPad showed scores were vulnerable to glare, but including anti-reflective screen covers and positioning to avoid reflections removed the effect [41]. Of note, later generations of tablet screens claim to incorporate anti-reflective coating [55]. Further study is merited to

evaluate if such changes in screen technology negate the need for such adaptations.

Aging of tablet screens and of fluorescent bulbs used to illuminate ETDRS charts affect luminance, and possibly luminance variability, and no attempt was made to control for this. An ETDRS chart manufacturer cites lamp life as 9000 hours [59], which equates to an approximate lifetime of 4.3 years of usage assuming use for 8 hours per day during a working week. They further advise annual lamp replacement when cabinets are used for research purposes. In the present study, ETDRS charts #2 and #3 exhibited the poorest luminance uniformity and had been in active clinical use for over 6 years without lamp replacement, falling outside the recommendations made for research purposes. The authors recognise this limitation, and it is likely that the aging of the bulbs has impacted on the results for these charts. However, the most uniform ETDRS chart (#1) was the newest, having been used for 2 years and less than 9000 hours in a clinical trials unit, meeting manufacturers recommendations [59]. Even this chart failed the BS 4274:2003 criterion for luminance uniformity when tested with the room lights off. All charts, however, easily exceeded mean luminance and contrast requirements, again highlighting that the BS 4274:2003 criterion for luminance uniformity may not be adequately evidence-based. All the iPad Air 2 devices were less than 6 months old, with variable usage. The oldest tablet was the iPad 3, which was over 3 years old and in daily use for both recreation and clinical testing: its parameters fell within the range of the other six devices, suggesting that age may not be a major factor in photometric compliance.

Another consideration when applying standards set for charts to tablet devices relates to the significant differences in size of the test area in both platforms. The angular field of view from a tablet device, together with the extent of the background relative to the optotypes, represents an intrinsic difference that could effect clinical results. In a recent study comparing a mobile phone-based single-optotype tumbling E test on the 4.8 inch

screen of the Samsung S3 handset (Samsung Electronics), with the ETDRS chart, 272 patients were assessed using both platforms [36]. An average difference of 0.011 logMAR was detected (95% limits of agreement -0.31 to 0.42), suggesting that the impact of overall test area is unlikely to be a clinically significant limitation.

Clinical relevance

The present study suggests that photometric standards of luminance, contrast, and luminance uniformity are met more effectively by the iPads under test (generation 3 to most recent model, the iPad Air 2), than by the dedicated ETDRS charts in active clinical use. The adjustable nature of the nominal brightness setting on such devices allows a relatively simple calibration process. The study fulfilled its aim of documenting the suitability of this mobile technology with reference to high contrast visual acuity test standards, and provides a practical guide to health care professionals working in this field. Specifically, setting the nominal brightness setting to 75% optimised the iPad with respect to ICO and British Standards recommendations (BS 4274-1:2003). These findings are naturally time-limited: due to rapidly changing technology, updated versions reach the market within around two years of the preceding model. Assertions that are valid for the latest platform may not be directly applicable to the next. Furthermore, device manufacturers control the legacy of the incorporated software and hardware without any necessity to conform to standards for chart design, which means it is possible for devices of the same name and generation to have different screen properties. Nevertheless, given that wide fluctuations in luminance result in relatively small changes in target contrast, and consequently small effects on acuity scores, the current data support the view that the increasing use of tablets and similar devices is not likely to be unsafe for clinical high contrast acuity measurements provided that simple checks, based on the adjustable brightness of each device, are conducted.

2.5 Conclusion

Healthcare is approaching an impasse where low-cost mobile devices designed for recreation will contain sophisticated assessment capability significantly in advance of those of hospitals and health centres. For vision, clinical standards represent a crucial reference to guide healthcare scientists and clinicians, but the ever-changing digital technology landscape challenges standard setters. Unregulated or inaccurate software/hardware combinations which purport to provide diagnostic information risk harm to patients yet failing to adopt the best of these technologies risks missing opportunities for novel or lower-cost techniques to detect visual impairment. Robust scientific validation, broadening the remit of international standards, and provision of practical guidance represent possible ways we can maximise the safe adoption of this ubiquitous technology.

Chapter 3: Evaluation of Conventional Card Based Acuity Tests and Digital Platforms

3.1 Introduction

The design and testing method of the described standard infant acuity tests vary significantly. However, all forced choice preferential looking paradigms depend crucially on the visibility of a given target relative to the otherwise featureless background, designed to match the average luminance of the foreground gratings or pseudo-high pass lines.

While stringent standards exist for adult acuity chart design, as detailed in Chapter 2, no equivalent standards exist for infant vision, with no guidance on expected lifespan of the test cards.

If the design of the test is adequate, it may be the case that trivial colour / grey mismatches make negligible impact in terms of a clinically meaningful behavioural response. However, to understand whether these variables bring potential for inadvertent visual cues, we first need to robustly measure them. Once measured, the potential to quantify and, if the case, recreate these deviations as independent variables then allows us to unpack their impact on detection thresholds.

It is our hypothesis that variations in handling of the (pseudo?) iso-luminant grey element of conventional forced-choice preferential looking tests is a significant factor contributing to the previously detailed observed differences in behavioural results observed between infant vision tests. For each of the reference standard infant vision tests, as well as a rage of digital displays, this chapter aims to detail how closely the actual background “grey” values match the iso-luminant target (based on the black and white foreground grating elements). In addition, we evaluate different digital techniques to attain an optimal iso-luminant backdrop.

The present study aims to develop the theme of the preceding Chapter, extending the standards applied to traditional (non-infant) optotype tests to the infant platforms under test, and report their compliance/deviation. For card-based tests dependent on reflected light, we repeat the analysis under 3 illimitation conditions: hot/cold studio light and fluorescent ambient light.

We also investigate different approaches to the digital generation of a grey background on a screen. In parallel research and development, as will be detailed in Chapter 4, PV deployed a black/white checkerboard at the maximum resolution of the display to achieve the target luminance [37]. In the present study, we evaluate alternative approaches to match the average luminance of the foreground black/white gratings, comparing a checkerboard pattern, a vertical grating pattern and a horizontal grating pattern, all at the maximum device resolution.

Desktop monitor, laptop and smartphone screens were added to the gamut of index digital platforms to evaluate whether the observations for the various computerised displays were consistent and establish their potential to mount a reliable infant vision test relative to existing standards for letter acuity tests. If the wider range of devices meet the required standards, or indeed exceed that of traditional cards, the potential to leverage web-based infant vision tests and extend the reach beyond a single proprietary tablet platform could be explored in future studies.

3.2 Methodology

In this experiment, two aspects were measured the luminance of the light patterns for traditional tests and electronic displays:

The traditional tests measured were:

Teller cards (Vistech Consultants Inc. Ohio, USA), Keeler cards (Keeler, Windsor,

UK), Lea paddles (Lea-Test Ltd and Good-Lite, Illinois, USA), Cardiff cards (Good-Lite, Illinois, USA). These platforms are demonstrated in Figure 1 (page 19)

The screens tested were: iPhone 6 (Apple Inc, California, USA), iPad 3 (Apple Inc, California, USA), Laptop screen (MSI GL62M 7RD; Micro-Star Int'l Co. Ltd., New Taipei, Taiwan), 4k HD monitor (Philips BDM4350; Koninklijke Philips N.V., Amsterdam, The Netherlands).

Five luminance and spectral measurement sets were taken in different locations to measure the difference between different points of the screen, similar to measuring techniques reported elsewhere [40]. With the digital displays, the five points were located as shown in Figure 5.

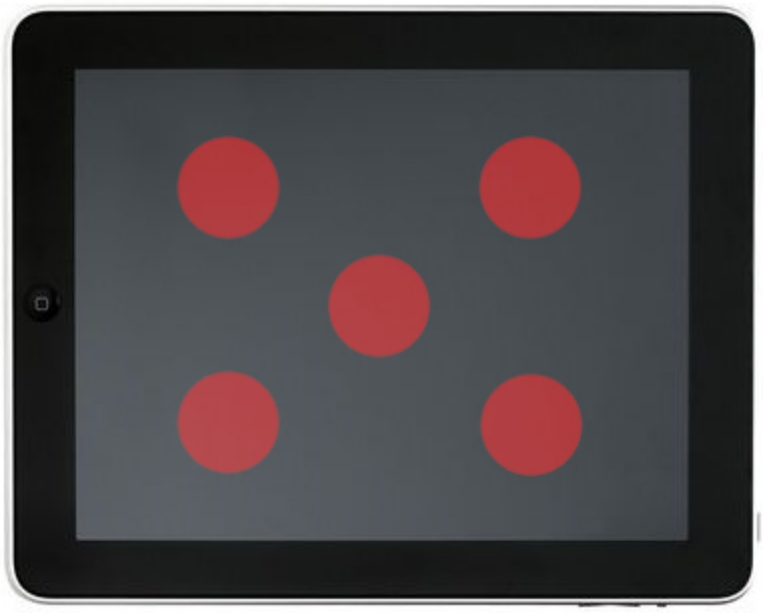


Figure 5. Location of the measurement points for the electronic displays.

Each point is located at a distance of 25% of the two nearest screen edges, except for the middle point which is located at 50% of distance of all the screen edges.

Because the traditional card-based tests had different target shapes due to the nature of the tests, the location of the five measure points was not the same across tests. The location can be seen in Figure 6.

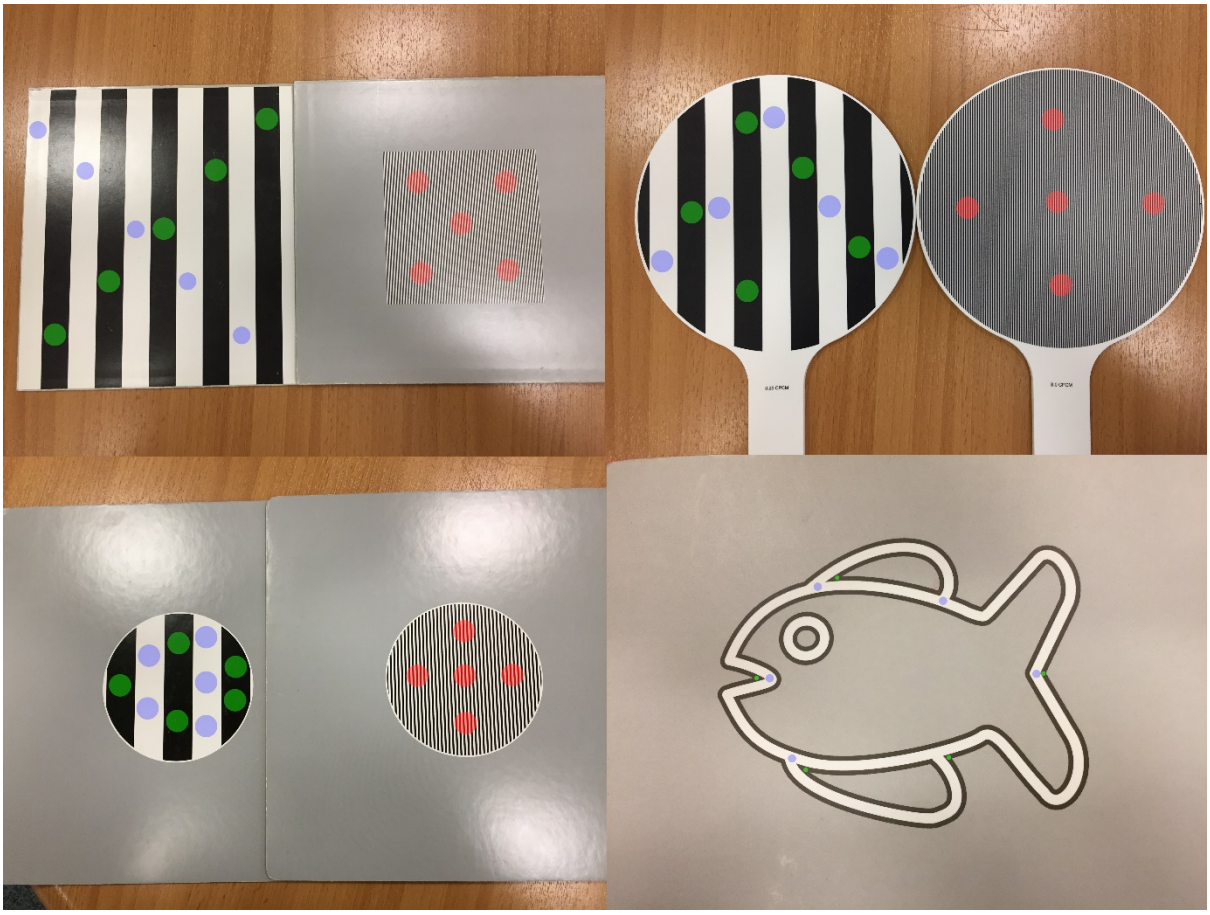


Figure 6. Location of the measurement points for the traditional tests.

Red dots indicate the location of the measurements for grating zones of targets; blue dots indicate the location of the measurements for white zones of targets; green dots indicate the location of the measurements for black zones of targets. For the measurement of the grating zones, the points were located at 25% of the edges of the local grating area, except for the middle point which was located on the centre of the grating area. Locations of the measurements for zones of black and white were chosen to distribute them as evenly as possible throughout the whole area of the tests. Note that the gratings used for this figure are not necessarily the finest gratings available, but rather thicker gratings to allow for better visualization in the figure.

The electronic displays were turned on for at least 15 minutes prior to the measurements [40] and set to a screen brightness of 50% with the auto-brightness feature of the devices turned off.

The measurement of luminance was performed using a luminance meter Minolta LS-100 (Konica Minolta Sensing, Europe B.V.) with a 1° aperture and a calibration traceable to the standards *Luminous Intensity Standard Lamp No. CS-0142* and *Master Diffusion Plate No. CS-0013*. The luminance meter was adjusted using a close-up lens No. 110.

The measurements were taken with the target placed on a horizontal surface and the luminance meter held perpendicularly, at a distance of 71±5 mm, positioned to minimize any shadows cast on the target from objects in the vicinity, the luminance meter itself, or the researchers while taking the measurements.

Luminance measurements were taken under three illumination conditions:

Ambient light from fluorescent bulbs (Sylvania CF-LE 40W, LEDVANCE, North Carolina, USA), similar to those found in clinical environments.

Light from two led studio lights with tuneable colour temperature (Aputure Amaran AL-H198C; Aputure, Shenzhen, China), located at an angle of 45° from the normal to the target surface, illuminating the target from two sides:

Cold light at 5500K (K: Kelvin), similar to daylight

Warm light at 3200K, similar to incandescent bulbs

No natural light affected the measurements because the room where the measurements were taken had no windows.

The light level was measured with a digital light meter (ISO-TECH ILM-01; RS Components, Corby, UK) at 0, 15 and 30 minutes after starting the readings. For the studio light conditions, the studio lights were left on 30 minutes prior to the measurements to allow the intensity to stabilise.

The light level (target illuminance) with the ambient lights on was, on average, 450 lux. With the studio lights on, the light levels were, on average, 3682 lux and 3432 lux for cold and warm light respectively, with variations of less than 1% of the average value over the course of the measurements.

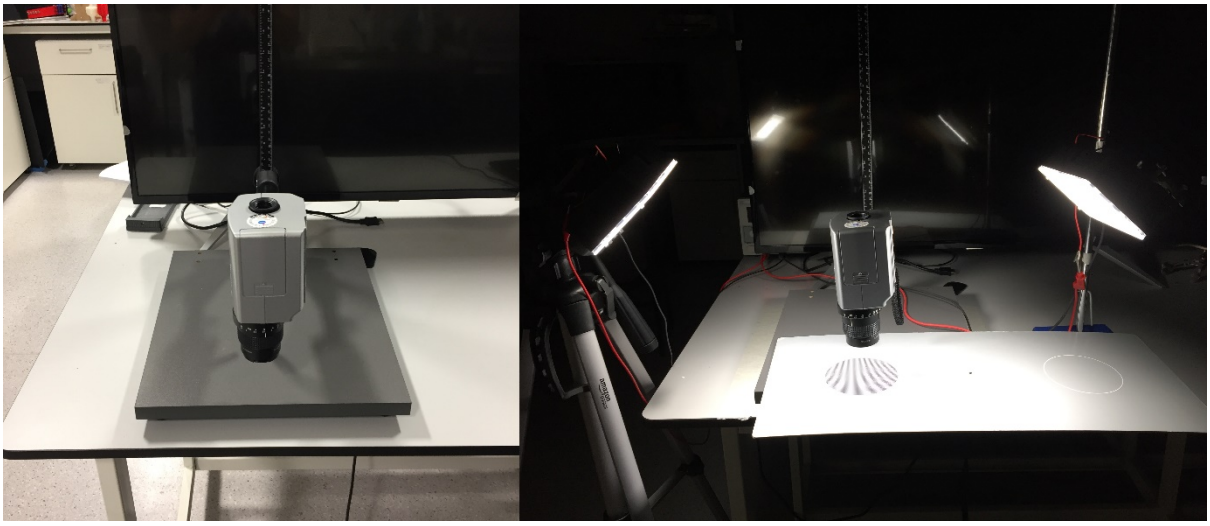


Figure 7. Layout for the luminance measurements under ambient light (left) and under studio lights (right).

For the traditional card/paddle tests, luminance measurements were taken of pure white and pure black from the card/paddle with the thickest grating available for each test; of grey background; and of the two finest gratings available, averaged over the field of view of the measurement instrument. There were two exceptions to this:

Cardiff cards only allowed to measure pure black, pure white and grey background: because they are based on the principle of a vanishing optotype, composed by 3 lines that are printed thinner for each increasing acuity testing level, measuring the smallest

gratings was not possible with our current equipment, due to the small size of the target details.

Lea paddles did not allow to measure the second finest grating with the close-up lens because the size of the grating was too high to allow a reliable measurement. For this target, measurements were taken without the close-up lens at 106 cm (± 1 cm) and normalised to the luminance of the close-up lens condition.

For the electronic displays, the measurements were taken for pure black, pure white, and three gratings made up of a checkerboard pattern, a vertical grating and a horizontal grating. The last three patterns were created using the OpenCV (computer vision) libraries [60], at the maximum resolution of each device – i.e. each half-cycle of the grating had 1 pixel of size – combining white pixels (RGB 255, 255, 255) and black pixels (RGB 0, 0, 0).

No measurements were taken of a purely grey background in the electronic devices because grey background can be simulated using fine grating patterns with pitch below any reasonable visual acuity resolution limit, such as those developed within related research and development (1, detailed in subsequent Chapter). Interestingly, in three out of four of the traditional tests (Teller, Keeler and Cardiff cards), the grey background is not printed as a pure grey, but as a pattern of black dots over a white background (see Figure 8).

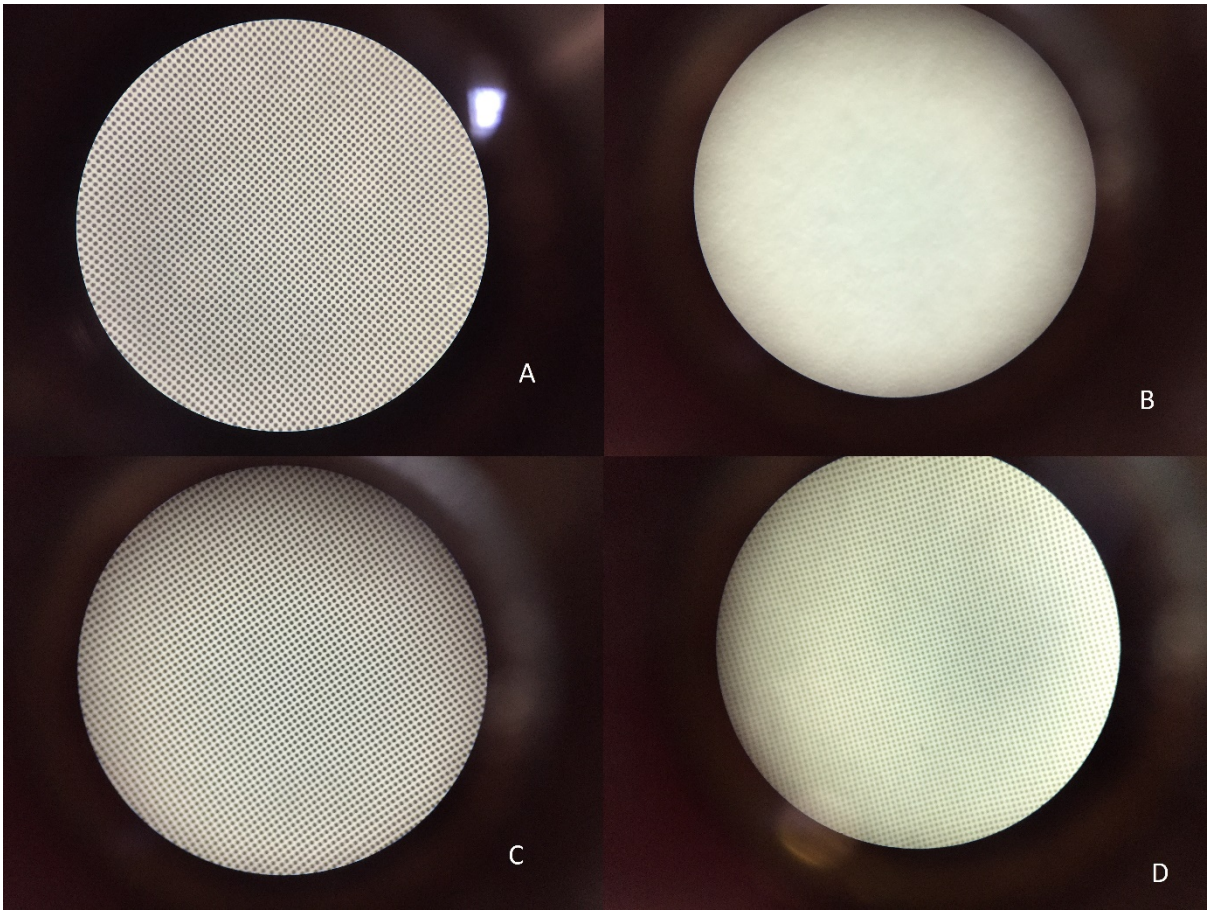


Figure 8. Close-up look of the grey background for the traditional tests.

A) Teller cards. B) Lea paddles. C) Keeler cards. D) Cardiff cards

Once the measurements were taken, on the traditional tests the luminance of grey background ($L_{background}$) and gratings ($L_{grating}$) were used to calculate Weber contrast (C_W)

$$C_W = \frac{L_{grating} - L_{background}}{L_{grating}}$$

Additionally, the maximum luminance of white ($L_{White\ max}$) and minimum luminance of white ($L_{White\ min}$) in each platform and condition was used to calculate the uniformity as

$$Uniformity = \frac{L_{White\ min}}{L_{White\ max}}$$

Because no specific standards for infant card/paddle-based acuity tests have been found, a comparison was made with ICO [50] and British Standards [51] for general adult acuity tests. The most relevant ones appear to be the ICO overall luminance standard (luminance > 80 cd/m^2) and the BS 4274–1:2003 overall luminance (luminance > 120 cd/m^2) and uniformity standards (luminance uniformity > 80%).

3.3 Results

3.3.1 Luminance

The results from the luminance measurements can be seen in Table 7. The measurement of grey value returned different values depending on the card measured for the Teller cards test, which can be seen in Table 2. No noticeable difference was noted for the luminance of grey in the different cards in the Keeler card range (also shown in Table 8.).

	Ambient light (450 lux)					Cold light (3682 lux)					Warm light (3432 lux)				
	Black	White	Grey	F.G.	S.F.G.	Black	White	Grey	F.G.	S.F.G.	Black	White	Grey	F.G.	S.F.G.
Teller cards	7.462	128.9	FG:67.3 SFG:59.42	67.57	56.04	56.43	1038	FG:528.3 SFG:471.58	541.8	450.3	51.74	949.5	FG:487.2 SFG:431	495.8	411.7
Keeler cards	3.132	127.5	FG :59.78 SFG :59.17	58.76	59.46	24.5	1028	FG:482.8 SFG:478.3	475.6	481.9	21.75	933.6	FG:482.8 SFG:478	433.7	428.7
Lea paddles	7.403	136.1	66.18	49.31	N.M.	52.22	1079	532.2	387.5	N.M.	47.94	1002	479.7	352.0	N.M.
Cardiff cards	21.12	124.8	69.40	N.M.	N.M.	161.8	984.0	545.4	N.M.	N.M.	154.5	905.7	500.7	N.M.	N.M.

Table 7. Averaged luminance (cd/m^2) for the card-based tests of pure black, pure white, grey background, finest grating (F.G.) and second finest grating (S.F.G.).

Measurements for the Cardiff cards were not possible to obtain for the finest and second-finest, as the design of the composite lines did not lend itself to measuring an average luminance measurement in the same way grating patches do. Poorly repeatable values precluded average measurements for the second finest grating in the Lea paddles (as detailed in methodology).

The luminance value of grey varied for each Teller card test, but not in the Keeler card range, hence entered as multiple entries. These results are further detailed in Table 8 (*).

Black areas showed the greatest relative difference across the traditional tests while white areas showed the lowest relative difference. This was consistent for all illuminating conditions. Apart from the differences, it was noted that, as reasonably expected, the room light level had a direct impact on the luminance measurements for for each aspect of each test

	Ambient light (450 lux)			Cold light (3682 lux)			Warm light (3432 lux)		
	Thickest grating card	Finest grating card	Second finest grating card	Thickest grating card	Finest grating card	Second finest grating card	Thickest grating card	Finest grating card	Second finest grating card
Teller cards	71.95	67.34	59.42	570.5	528.3	471.58	519.3	487.2	431.06
Keeler cards	59.81	59.78	59.17	481.86	482.8	478.3	428.7	442.0	434.7

Table 8. Averaged luminance (cd/m^2) of the grey background in the Teller cards for the three cards measured in the experiment.

The results from the luminance of the electronic displays are shown in Table 9.

	Ambient light on (450 lux)					Ambient light off				
	Black	White	H.G.	V.G.	C.G.	Black	White	H.G.	V.G.	C.G.
Phone	1.036	149.2	74.36	76.46	73.34	0.108	155.8	76.73	77.69	75.88
Tablet	1.242	130.4	65.49	65.79	65.53	0.151	127.4	64.09	64.22	63.64
4K display	1.502	168.3	81.92	83.22	81.46	0.158	169.5	81.41	83.75	81.82
Laptop	2.637	123.6	64.51	64.08	63.26	0.683	124.5	62.93	61.50	62.49

Table 9. Averaged luminance (cd/m2) for the electronic displays of pure black, pure white, horizontal grating (H.G.), vertical grating (V.G.), and chessboard grating (C.G.)

For the screens, turning the lights off did not produce a noticeable difference in the luminance, except for the luminance of the black areas. Additionally, all three grating patterns presented similar luminance values within each device.

Under each tested light condition, all the card-based tests and the electronic devices met the criteria for both the ICO (International Council of Ophthalmology, 2011) overall luminance requirement (>80 cd/m^2) and the BS 4274–1:2003 overall luminance requirement (>120 cd/m^2).

3.3.2 Contrast

The results of the contrast between the gratings and the grey background can be found in Table 10. In the case of the Teller cards, because the grey background luminance was found to be vary widely between cards, the contrast was calculated between the grating and the grey background of the card containing that grating.

	Ambient light (450 lux)		Cold light (3682 lux)		Warm light (3432 lux)	
	Finest grating	**Second finest grating**	**Finest grating**	**Second finest grating**	**Finest grating**	**Second finest grating**
Teller cards	0.3	-5.7	2.6	-4.5	1.8	-4.5
Keeler cards	-1.7	0.5	-1.5	0.8	-1.9	-1.4
Lea paddles	-25.5	-12.2	-27.2	-12.0	-26.6	-10.3

Table 10. Weber contrast (% difference) of gratings vs grey background for Teller cards, Keeler cards, and Lea paddles. With Teller cards, the contrast was calculated between the grating and the grey background of the respective card.

The contrast between the gratings and the grey background was greatest with Lea paddles, presenting a marked deviation from target average luminance, relative to averaged grating luminance.

Figure 9 plots the results from the luminance and contrast measurements, where the upper panels represent the traditional tests and the lower panels represent the electronic displays. This demonstrates a notably wider divergence between target average luminance of background, and also a wider divergence of the finest grating from the target background luminance. While generally a more notable issue with the traditional platforms (upper panel), it appears most pronounced for Lea paddles. In addition to this, the luminance of the finest gratings is particularly distracted from the target luminance level (closed circles, Figure 9, upper panel).

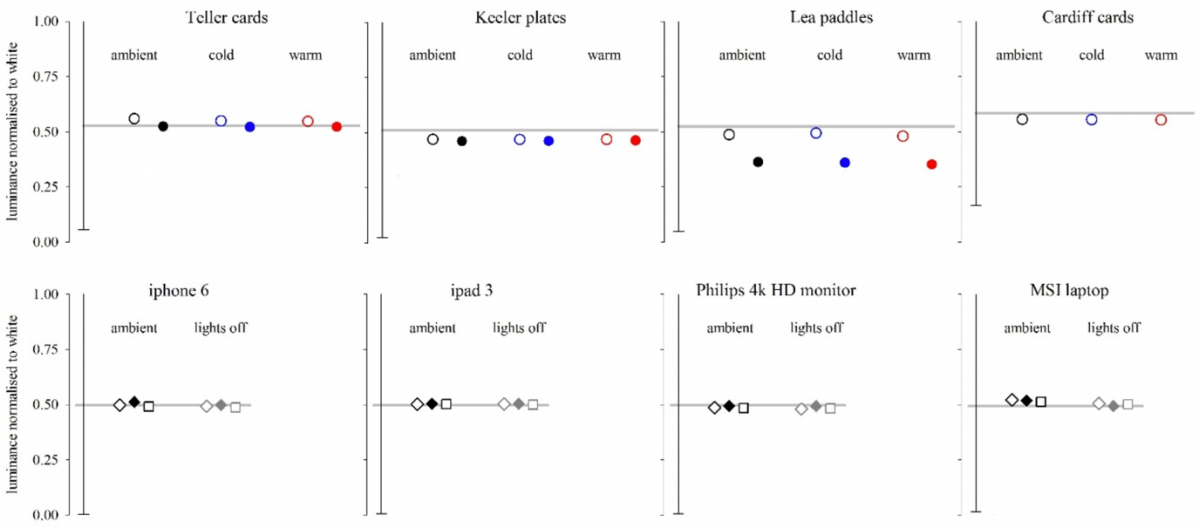


Figure 9. Relative luminance of target white and black zones, of grey backgrounds, and/or grating zones for the four traditional tests (upper panels) and the four electronic displays (lower panels).

The luminance of black and white are plotted on a vertical plane (normalised to the luminance value of white for each platform), together with the theoretical luminance of the ideal grey of the grey background (horizontal line) and the actual luminance values of the grey background and the gratings (symbols).

Upper panels: open circles represent grey background; closed circles represent the finest grating.

Lower panels: open diamonds represent horizontal grating; closed diamonds represent vertical grating; squares represent chessboard grating. Illumination condition is colour-coded: for the upper panels ambient fluorescent lighting in black, cold light in blue, and warm light in red; for the lower panels ambient light on in black and ambient light off in grey.

3.3.3 Uniformity

The uniformity of each platforms can be found in Table 5. All platforms were compliant with the BS 4274–1:2003 uniformity standard (uniformity > 80%) in all the illumination conditions.

	Ambient light	Cold light	Warm light	Lights off
Teller cards	97.99	97.61	98.54	N.M.
Keeler cards	99.53	99.41	98.97	N.M.
Lea paddles	99.41	97.79	98.94	N.M.
Cardiff cards	97.70	97.50	97.36	N.M.
Phone	91.28	N.M.	N.M.	90.82
Tablet	90.17	N.M.	N.M.	89.05
4K display	93.08	N.M.	N.M.	94.00
Laptop	93.45	N.M.	N.M.	88.5

Table 11. Uniformity of the platforms under the different illumination conditions (%).

Luminance measurements were not taken for the card-based tests with the lights off, nor for the electronic displays with the studio lights (N.M.).

In the case of the traditional tests, uniformity was above 97% for all tests under all illumination conditions. Illuminating under studio lights reduced uniformity in three of the tests (Keeler cards, Cardiff cards, Lea paddles) and increased or decreased depending on the colour temperature of the studio light in one test (Teller cards). The variations in uniformity were lower than 1% for all traditional card/panel tests under the different illumination conditions.

The electronic displays all showed uniformity levels above 90% with the ambient light

on. When the ambient light was turned off, three out of four of the devices decreased their uniformity (the 4k display was the only one where uniformity increased). The most marked decrease was for the laptop, with a drop of approximately 5%, whereas all the other devices only showed variations of around 1% with ambient light on and off.

3.4 Discussion

While there is a strengthening evidence base regarding the leveraging of ubiquitous mobile technologies for children's vision testing with rapid evolution of digital display technologies [38–40,62,63], the reference standard for infant vision is nevertheless essentially unchanged over the past 35 years [14,15,45]: a printed surface. This reference standard, critically tied to the practical limitations of the ephemeral nature of the paper and ink used in the printing process, appears in complete antithesis to the fast pace of technological advance in almost every other field of ophthalmic diagnostics.

The barriers to traction of available digital tests are multifactorial. With ubiquitous internet-based repositories for digital applications purporting to check vision, the clinician cannot be certain regards which platforms/software to trust (Perera et al, 2015). While one of the largest trials for large-scale paediatric screening centred around a specific mobile device brand and model [36]. There are varying approaches from mobile handset manufacturers regards pixel-density handling, which has been recognised to lead to inaccurate optotype sizing [64], and hence a significant potential for inaccurate clinical information exists, limiting scalability. For those applications that do appear to have been robustly evaluated, the short life-cycle of the specific tablet/smartphone platforms, together with the wide gamut of device ranges/generations, engenders a particularly elusive platform stability for medical device developers.

While CE and FDA marking appears to lend confidence to the end users, regulatory authorities fail to capture the photometric pre-requisites necessary for such platforms to

diffuse into routine practice. The present study, in line with a growing number of similar studies [38–40,63] demonstrate innate advantages to Lambertian retro-illuminated displays for visual testing purposes, as well as workarounds to compensate with gamma correction to maximise efficacy and range for contrast testing [38].

Interestingly, the rigorous evaluations applied thus far to mobile device screens for the purposes of vision testing have *not* been applied, to our knowledge, to the recognised gold standards for infant vision testing. Indeed, at the time of writing, the present group can find no national or international standards relevant to the physical and/or photometric properties of leading card-based tests. Given the central premise of force-choice preferential looking (FCPL) paradigms relates to the background grey being of equal average luminance to the test gratings [13,65], ensuring the grey is standardised is vital to performance. Although differing manufacturers have approached generation of the background grey in different ways (as demonstrated in Figure 8), no standard exists to measure against, and the impact of card/printed surfaces degradation/fading over time is unclear. Deviation from an equal average background luminance raises questions on the origin of just-noticeable differences within a card/between paddles, giving rise to a potential for a misrecognised conflation of contrast and grating acuity measurements.

3.4.1 Luminance

The luminance of the test areas (black, white, grey and gratings) presented differences across the traditional tests, although the differences were not constant for all the areas. In relative terms, the black area presented the highest difference across tests while the white area presented the lowest. The reason for these differences may be due to the lack of harmonized standards for the traditional tests and different approaches in the manufacturing process regarding the materials used for the base and the printing. However, this does not necessarily imply clinical significance: because of the lack published data, it is not possible at this moment to say if these differences in luminance

translate into difference in performance in the clinical field, which is a focus of future research.

An interesting finding during this experiment was that with Teller cards, the luminance value of the grey area was not consistent across cards. The value of the luminance of the grey background differed across the thickest grating, the finest grating and the second finest grating. While the reason for this difference is not known as of now, it is theorised that the aging of the cards may have degraded the printed colour, hence resulting in the observed difference. This difference in grey backgrounds across different cards was not found in the Keeler cards, which had a more consistent luminance value of the grey area within the tested platforms.

In the case of the electronic displays, there were also differences across the platforms for all test areas. However, electronic displays can be calibrated in terms of luminance and contrast to match other devices and any requirements of national and international standards, or of specific applications. Furthermore, when normalised to the white value, luminance levels across the electronic devices show very little difference, as demonstrated in Figure 9. In general, the black area in the digital devices had a lower relative value with respect to the white area than in the traditional tests, and the gratings of the electronic devices better matched the target theoretical luminance (average between black and white luminance) than the gratings of the traditional tests (Figure 9).

Another important feature noted in the electronic devices is that the luminance appeared to be barely affected by the illumination conditions. An exception to this was the luminance of black areas which showed a significant relative decrease when the ambient light was turned off, although in absolute terms the increase was rather low (in the order of 1-2 cd/m^2). This is thought to be because in a black screen none of the pixels are emitting light, hence the reflectivity of the display accounts for most of the luminance

the value measured.

Regarding the standards from the ICO (International Council of Ophthalmology, 2011) and the BS (British Standards Institute, 2003), all the measured platforms fulfilled the requirements of both. Predictably, the luminance of the traditional tests were directly related to the illuminance of the room. From this, it can be deduced that the illumination condition of the room plays an essential role in the compliance with the overall luminance requirements. The apparent lack of effect of the ambient illumination conditions over the luminance of the electronic displays offers an advantage over traditional tests as the environmental conditions may not need to be as regulated as in the case of traditional tests.

From a luminance point of view, electronic display technologies appear to have advantages on the core physical luminance properties over the traditional tests. Electronic displays can be adjusted to match any luminance requirements and less effected by the illumination conditions of the room. Additionally, they appear to be more robust to degradation over time and it is possible to re-calibrate them if the quality of the image degrades. It should be noted, however, that while the tablet device tested is over 7 years old at the time of writing, the paddles and cards have been in clinical use for longer, and while we do not have clear records on age of cards, the Teller and Keeler cards have been in clinical use for > 10 years. The performance of digital platforms over such an extended timescale is uncertain, and this bias may have impacted on results.

Furthermore, the present research did not evaluate the impact of screen reflections, which is likely to be an issue the digital displays are more prone to, particularly in brighter ambient conditions.

3.4.2 Contrast

The luminance matching between the gratings and grey background is an underpinning assumption of the test paradigm. An unintentional contrast difference between grating and background could result in a child discerning the gratings over the grey background due to contrast thresholds, conflating the result, which is mistakenly ascribed a high contrast acuity score.

This is considered by us to be one of the key aspects of the photometric compliance of these tests, and the results raise concern regards the absence of a regulatory standards.

The results show that the tests do not handle this contrast in the same way, with Lea paddles having the most divergent under all illumination conditions tested. If these results were correct, it would indicate that Lea paddles present an important photometric deficiency that could result in a decreased effectiveness of the test with respect to the other tests. The other tests also present luminance discrepancies between the grey background and the gratings, although to a significantly lesser extent. Importantly, however, we note that the clinical relevance of these photometric findings is yet unknown due to the lack of published data on the matter. Anecdotally, some of these contrast differences could be discerned at long distances, generating an acuity score significantly better than accepted norms, with members of the authorship identifying the finest grating paddle at protracted distances up to 10 meters. However, contrast sensitivity develops with age, and hence, infants may well react differently [66].

Using the studio lights, mimicking daylight and incandescent light, as a source of illumination did not appear to have a clear effect on contrast. While the contrast measured was different under the different illumination conditions, the differences did not show a common trend across tests (in some cases it increased and in some other cases it decreased) and, additionally, the differences in contrast were relatively small, in

the order of 1 – 2%.

Contrast appears to be one of the most compelling arguments in favour of the digital device platforms. Using the checkerboard grating technique to generate grey backgrounds (described in 1) ensures, on a fundamental basis, the luminance of the grating targets and the grey background is the same, eliminating a possible defect in the test due to contrast. However, as with luminance values, clinical validation is required to understand if this intrinsic lack of contrast is a clinically meaningful advantage.

3.4.3 Uniformity

All platforms were compliant with the BS 4274–1:2003 uniformity standard (> 80%) under all illumination conditions, which did not appear to play a major role in the uniformity of any of the platforms.

The traditional tests showed variations of less than 1% in uniformity when exposed to different illumination sources. Given the uniformity values for the traditional tests were above 97%, this variation may be considered negligible.

The electronic platforms showed a similar trend with lower general uniformity (values between 90 – 94% approximately). The exception was the laptop which showed a decrease in uniformity of approximately 5%, though the platform was still above the 80% uniformity requirement.

In terms of uniformity, traditional tests appear to present advantages over the digital tests, both in overall value of uniformity and difference when changing illumination conditions.

3.4.4 Study limitations

The photometer used for this experiment was used at a close distance with a close-up lens. The magnifying effect of the close-up lens made it so that, in the measurements of

the gratings, the number of cycles captured by the measuring area of the photometer was low, hence potentially resulting in luminance measurements that deviate from the real values.

Although the cards were deemed generally in good condition, as the cards were in active use across various hospitals and academic units, we could not control for the aging and storage of the traditional tests or the displays prior to the measurements of this study. The aging and storage conditions of traditional tests may be an important factor in their performance. It is quite possible that the Teller and Keeler card collections represent a heterogenous collection of different and unknown ages. How an iPad, laptop or computer monitor is likely to perform at an equivalent age (cards likely to be 10+ years in active clinical use) is an open question. Indeed, cards are prone to wear and degradation over time. Yet again, this aspect is not regulated by any standard. In the case of the electronic displays, even without considering the aging, all complied with the standards of luminance and uniformity used for comparison in this study.

While every effort was made to avoid obstructing light illuminating the platforms during the measurements, it is recognised that in some cases this might have happened to a minor degree. The very positioning of the researcher during the measurements may have caused an impact in the illumination affecting the card, which was not possible to account for.

3.5 Conclusion

Taken together, findings from the present study demonstrates intrinsic advantages to digital platforms regards achieving a mean average Luminant “grey” background relative to the black and white values of composite foreground gratings in relation to high-contrast acuity testing.

There is significant heterogeneity across traditional card-based platforms regards their

physical properties in relation to FCPL high contrast acuity paradigms, which may go some way to explain why the norms vary widely across platforms [30–33]. Regulatory guidance in this area would go some way to harmonise practice and provide a framework around a minimum standard, allowing the high standards that exist for adult testing (ETDRS) to be extended to the infant population, where it can be argued the consequences of inaccurate measurements have greater consequences given the time-limited critical period of visual development unique to a paediatric population.

While the results show a surprising divergence between target and actual values in reference-standard tests, this is a laboratory-based study, and the clinical implications are unknown, so the relevance is presently unclear. It is ultimately the large normative data sets for Teller [67] and Lea [68,69] that make them trusted resources.

Ideally, subsequent research should capture the psychophysiological impact of the measured differences. To fully evaluate the clinical impact, further studies should pitch a conventional grating standard with a digital one, employing a similar paradigm to grey generation as used within the index test, peekaboo vision (Scottish Health Innovations ltd, Clydebank, UK).

Future studies should also seek to evaluate other factors that could introduce inadvertent cues that conflate acuity results, specifically evaluating the colour dominants introduced in the generation of composite black, white, grey and grating elements. Evaluating such factors across traditional and emerging digital platforms could inform and enhance design towards a more accurate and precise test basis.

Chapter 4: Evaluation of a Novel Digital Grating Acuity Test in an Adult Cohort

4.1 Introduction

Chapters 2 interrogated a tablet computer's display properties with reference to the National and International standards for visual acuity chart design. A similar evaluation has been extended to traditional card-based platforms in Chapter 3, with comparative analysis provided compelling evidence that digital displays offer advantages in terms of acuity test design and display, achieving a target neutral average luminance closer to the target mid-level of high contrast gratings – a key feature of FCPL paradigms,

In this chapter, we outline a novel paediatric acuity test design, and report our results in an adult cohort, comparing a novel paediatric acuity game with traditional standards.

4.2 Aims

A. Produce an infant acuity test for the Apple iPad 3 (Apple inc, Cupertino, California, USA)

Create a digital acuity game for pre-verbal children, which will offer preferential-looking functionality for infants and multiply disabled patients, lacking speech and coordinated movement. Allow functionality for older children, enabling independent operation by virtue of a touchscreen game format, without requirement of direct instruction by a specialist

B. Compare the tablet-based test, Peekaboo Vision, Build 1 (PVb1), with current clinical standards in an adult cohort.

While the Peekaboo acuity test is ultimately aimed at infants, testing in an adult cohort allows collection of preliminary data from a reliable population, less prone to loss of concentration when subject to multiple vision tests. Moreover, testing adults permits a more expansive range of acuities by artificially degrading vision (blurring spherical

lenses), which would not be tolerated in an infant trial.

The adult cohort allows an opportunity to refine the testing model and provide a measure of reliability in advance of testing young children as part of a pragmatic trial, evaluating diagnostic accuracy in primary care and pre-school screening.

4.3 Methodology

4.3.1 A. Digital Test Design

All graphical elements were designed and scaled using Adobe Photoshop Creative Cloud (Adobe Systems inc, San Jose, California, USA). The acuity test application was designed in HTML5, to the screen specification of the 3rd Generation iPad with Retina Display.

High-spatial frequency grating visual targets were chosen as the FCPL target. This method has been well described [16], and remains the clinical standard for infants in card-based preferential- looking tests, such as the Keeler Acuity Test for Infants (KACI, Keeler ltd, Winsor, United Kingdom), against which PVb1 was compared in the present study. KACI was selected as reference standard as it was the most commonly used FCPL grating test in our tertiary referral centre and felt to be represent the most typical UK test. The ideal test involves an engaging target that directs and holds the child's attention. In contrast to the traditional circular/square optotype targets, the digital high contrast grating in PVb1 targets were constructed as simple smiley-face graphics (Figure 10A). This format was chosen as faces have been demonstrated as salient to recognition in young infants, with newborns being attracted to faces from hours after birth [70]. It has been proposed that face recognition represents a special stimulus category, processed differently from other stimuli [71].

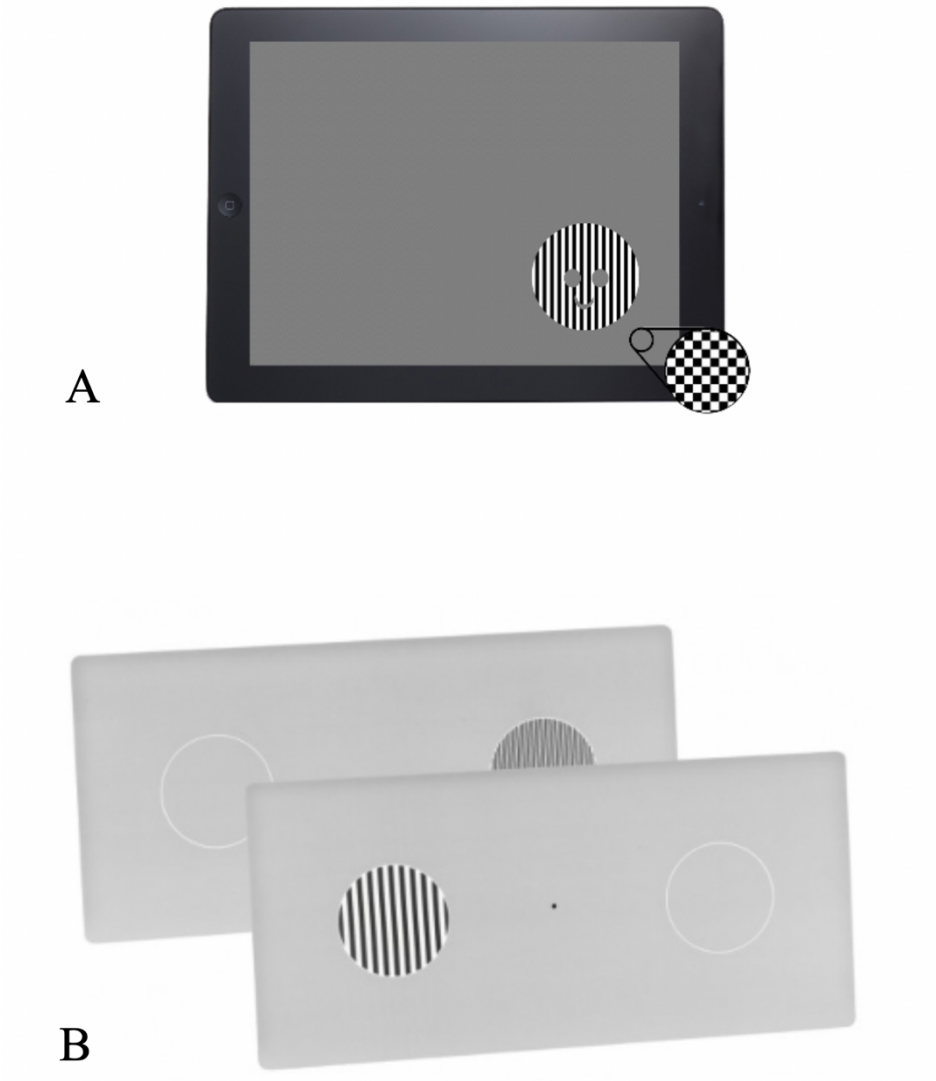


Figure 10 A: The Peekaboo Digital Acuity Test B: The Keeler Acuity Test.

To achieve a consistent and equal average luminance of the optotype target and background, in departure from card-based techniques that employ a homogenous grey, the background comprises an alternating black/white checkerboard pattern at the maximum resolution permitted by the device display (Figure 10A). To the observer, this appears as a uniform grey. This technique was demonstrated to precisely achieve the desired average luminance of the black/white foreground elements, as described in Figure 9 (lower panel, second from left, “iPad 3”).

The digital test was designed as a basic computer game, to be performed at 38cm. At this distance, the screen resolution is sufficient to assess grating acuities from -0.1 – 2.3 logMAR (equivalent to near blindness). The child is encouraged to tap or point to the smiling face graphic, following a demonstration by the tester with a low- spatial frequency optotype example. The optotypes are positioned consistently at 1 of 4 corners of the screen in a pseudo-random order. To prevent cues from the tester, the device is held facing the child in a position such that the tester is masked to the screen. For infants too young to point or press, an alternative technique is employed, whereby the tester infers which screen- corner the child is looking, by observation of the child's eye-movements. The tester presses the relevant quarter of the screen with the fingertips overlapping the edge of the device. If there are no meaningfully directed eye movements, the presented level of acuity is assumed to be below the level of detection, and lower- spatial frequency target is presented. A sound and animation reward is played if the position is correct. This sound alert includes a voice descriptor signposting the level of acuity attained, informing the tester without the need to rotate the device to assess where the graphic has appeared. This technique prevents disruption of the test sequence, and helps maintain a consistent working distance by limiting movement of the device.

4.3.2 Clinical Testing

Volunteers were recruited from departmental staff from Gartnavel General Hospital, Glasgow. All volunteers underwent reduced Snellen acuity testing (Sheridan Gardiners Near Acuity, Keeler ltd, Windsor, United Kingdom) performed at 33cm, with habitual correction if required. Eyes with near acuity poorer than 6/9 were excluded. The adult group (N=10, 20 eyes), underwent each acuity test monocularly with usual correction, if worn, plus 3 spherical blur conditions (+16, +8, +4 dioptre), which were placed over the habitual near correction if required. This allowed a total of 80 tests across a wide range

of acuities (-0.1 to 1.8 logMAR). All acuity scores were converted to logMAR for statistical analysis.

For all blur conditions, 5 minutes of blur-adaptation was provided prior to commencing testing. Acuity was measured with 3 tests: Reduced Snellen (33cm), KAC (38cm), and PVb1 (38cm).

The order of testing was randomised and equally balanced. To assess reliability, all tests were repeated 24-72 hours later.

For the digital Peekaboo acuity test, auto-brightness was set to "off", and the brightness set to 50%. In accordance with findings relating to iPad screen stability [38], the device was switched on, and display left to stabilise for 15 minutes prior to commencing clinical testing.

For both the KACI and PVb1, grating acuity was presented twice at each level, starting with lowest spatial frequency gratings. Testing continued in steps until an error was made, or the finest grating was correctly identified on two consecutive presentations. A forced- choice preference assessment model was adopted. The *staircasing paradigm* (the protocol deployed to guide whether finer or courser gratings are presented based on a given response) employed in this study is illustrated in Figure 15A.

AL (senior orthoptist) or IL (ophthalmologist) performed testing. The same tester performed test/retest for a given volunteer.

The variability of collected data was explored to assess reliability via Bland Altman analysis. For test-retest results, the inter-test differences were normally distributed.

4.4 Results

High-contrast grating patterns were presented at an equal distance in both test modalities. Theoretically, the matched line-widths subtend equivalent angles at the retina to provide equivalent results. However, results differed significantly between the card-based KACI and tablet-based PVb1 groups ($p<0.001$, paired t test).

Despite this, Bland Altman analysis comparing KACI with PVB1 results (Figure 11) revealed the mean difference in acuity results between the two tests was small, at less than one logMAR level (-0.07).

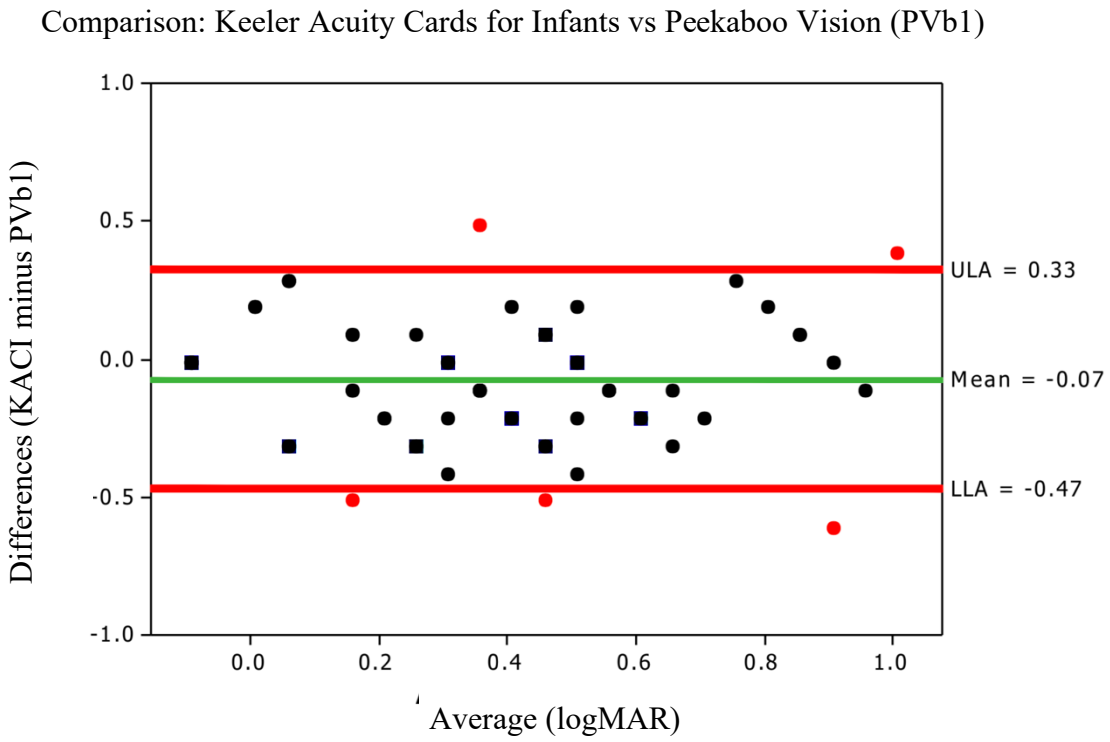


Figure 11. Bland Altman Plot demonstrating the difference between the Keeler Acuity Test (KACI) and Peekaboo Vision Protype Build 1 (PVb1).

Squares represent overlapping points.

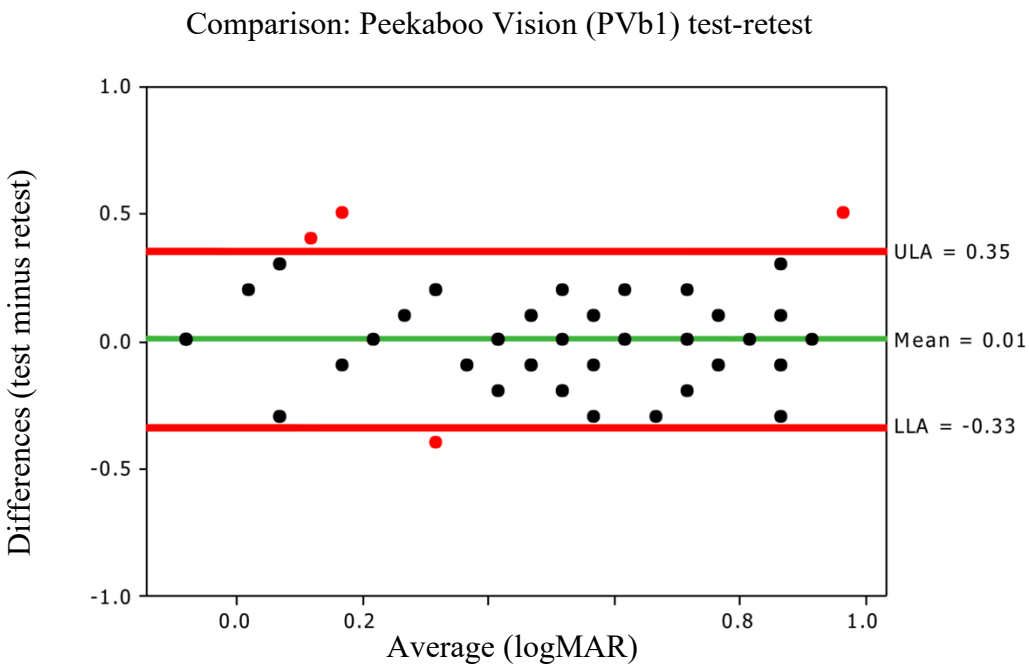


Figure 12. Bland Altman Plot demonstrating the difference between Test 1 and Test 2 (Retest) for the Peekaboo Vision (Build 1, PVb1), versus the mean acuity threshold.

The limits of agreement fall within approximately 1 octave.

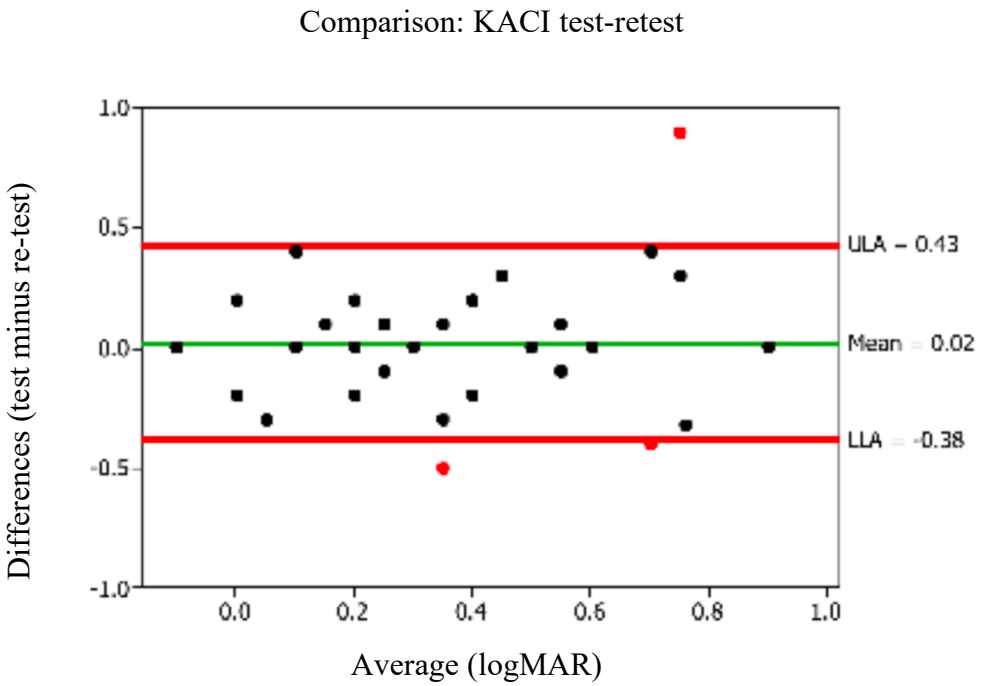


Figure 13. Bland Altman Plot demonstrating the difference between Test 1 and Test 2 (Retest) for the Keeler Acuity Test, versus the mean acuity threshold.

The limits of agreement are wider than that of Peekaboo Digital Acuity. Squares represent overlapping points.

On comparison of test-retest data, Figure 12 and Figure 13 illustrate superior reliability of PVb1, with 95% of the measured inter-test differences falling within 0.3 logMAR for PVb1, compared with >0.4 logMAR difference with KACI. In each difference plot, there is an even spread of detected differences, without weighting towards better or poorer acuities.

4.5 Discussion

Despite similarly sized targets being presented in both digital and card-based tests, the acuity results obtained with each test are significantly different (p=0.001), with the mean difference in results between KACI and PVb1 approximate to a difference approximate to 1 logMAR line (-0.07), with slightly better acuity (lower average logMARscore) being measured with KACI. This is likely to relate to the fundamental differences in physical characteristics of the device screen compared to card.

Limitations of the device include reflections and glare from the glossy surface. In addition, overall area of the optotype graphic varies, and the face graphic as an alternative to the circular gratings patch.

However, the tablet-based test evidenced improved test-retest performance on Bland Altman analysis. Future trials involving infants are indicated to definitively assess the value of this technology, and to address important questions relating to engagement with the test and time-to-test. If found to be a credible alternative to current clinical standards, app-based acuity tests could prove a disruptive technology in the field of infant acuity testing, with a potential role in amblyopia screening.

Particular potential impact in low- and middle-income settings. In 2000, 4% of the population in low- and middle-income countries had access to mobile technology. 15 years later, this has estimated to increase to 94%. In 2015, sub-Saharan Africa mobile cellular subscription is estimated at 76 for every 100 people [72].

The evidence base for mobile technology displays being well suited to acuity testing is growing [36,38,39,41,63,73–79]. With intelligence within software it is possible that with future developments in computer vision and eye tracking [75,76,80], poor vision could be

objectively detected with a digital solution leveraging the native front-facing camera [81], bringing potential to extend the reach even further, with dependence on a trained individual potentially obviated.

4.6 Conclusion

The present study demonstrates proof of concept regards a grating acuity test being delivered by an increasingly ubiquitous digital recreational platform in an adult cohort. With spherical lens blurring conditions, the prototype application, PVb1 appears to perform well across a range of artificially degraded acuities.

Next steps involve clinical trials comparing conventional standards with the digital platform in a paediatric population. The relative low-cost of the tablet platform, when compared with presently available card-based standards, brings potential promise that digital screening tests can be mobilised on a range of tablet devices, in turn bringing potential to extend the reach of reference standards to settings lacking access to prohibitively expensive traditional infant acuity resources. Future studies should also aim to evaluate performance in low- and middle-income settings.

Chapter 5: Testing Paediatric Acuity with an iPad: Validation of 'Peekaboo Vision' in Malawi and the UK

5.1 Introduction

Detailed in chapters 1-4, the advent of tablet computers creates the opportunity to test acuity digitally.

Computer tablets cost less than card-based acuity tests and can emulate vision tests as games. Automated stair-casing and reporting require fewer and less-specialised testers and could extend the reach of visual screening programs in both developed nations and in economies with limited access to healthcare.

The present study evaluates a digital, tablet-based forced-choice acuity test, 'Peekaboo Vision', designed to function as a preferential looking or touchscreen game, where the child can interact directly with the screen. We evaluate the technology in two paediatric populations, extending research and development detailed in Chapters 2-4, further evaluating a prototype high spatial frequency grating software, Peekaboo Vision Build 1 (PVb1), following it's proof of concept in a blurred adult cohort (1, Chapter 4).

PVb1 was first assessed in a cohort in Blantyre, Malawi, comprising Study 1. This pilot study allowed the concept of iPad-based digital gratings to be tested against a known reference standard and was instructive regards the subsequent formal software development. Study 1 also served as a test of methodology, guiding the design of study 2, where the formal build with a similar paradigm was evaluated in a UK cohort.

Differences between the heterogenous test populations, together with multiple intrinsic divergences in app builds (target design, staircasing, testable range and distance) make direct comparison between Studies 1 and 2 somewhat spurious. Data are hence presented separately for each study.

5.2 Methodology

The research followed the tenets of the Declaration of Helsinki. Informed consent was obtained from the subjects' parents/guardian after explanation of the nature and possible consequences of the study

5.2.1 Study 1

5.2.1.1 Patients

The Malawi College of Medicine Research and Ethics Committee granted ethical approval. 58 consecutive, unselected children, aged 6–60 months, presenting to the Lions Sight First Eye Unit in Blantyre were recruited, many of whom exhibited visual problems.

5.2.1.2 Setting

Testing was performed in clinic rooms by an ophthalmologist (LB or EM) experienced in testing paediatric vision. Testers were masked to any documented clinical information other than the age and date of birth of the child. This included diagnosis, past medical history and any previous visual measurements such as acuity, visual field, refraction or orthoptic problems. The same tester performed all tests on each child. Inter-tester variability was not investigated. All testing was performed in well-lit clinical areas by clinical staff with experience in testing children's vision. As the focus of the validation was to evaluate performance in a real-life environment against the present reference standard, no luminance measurements were taken regarding ambient light levels, or luminance levels from the Keeler cards.

All children were tested with both Peekaboo Vision and Keeler cards in random order. If the *day*-of-month within the date of birth of the child had an even number, Keeler was performed first. The both-eyes-open (BEO) condition was undertaken first. If the child had an even *month* of birth, then right eye was tested next, otherwise it was left first.

Typical occlusive glasses designed for paediatric acuity testing [82] were used to cover the fellow eye in the RE and LE tests. A handheld monocular occluder was also used for younger children who did not tolerate occlusive spectacles.

5.2.1.3 Peekaboo Vision build 1 (PVb1)

Study 1 used PVb1, as previously described. Graphics used Adobe Photoshop Creative Cloud and the application was developed in HTML5, to screen specification of 3rd generation iPad with 'retina display', (resolution 264 pixels per inch). Vertical grating targets were employed, being the most robustly described clinical standard for infants [67]. To engage and hold attention, grating targets comprised simple smiley faces (Figure 14 A) against an isoluminant background generated by an alternating black/white checkerboard at the maximum resolution of the display, which appeared as uniform grey. The detail of the eyes/mouth to comprise the smiley face elements were composed of the same isoluminant checkerboard pattern used to generate the background, such that they would only be visible if the grating could be delineated from the background. Test distance was 25cm, with acuity range of 0.1 to 2.2 logMAR (Table 12). Screen brightness was manually set to 50%. Measured acuity thresholds were expressed in logMAR.

The tablet was held in landscape orientation facing the child, so the tester was masked to target location. Targets appeared pseudo-randomly in one quarter of the screen following a 0.3 logMAR down, 0.1 logMAR up staircase (Figure 2A). Infants' eye-movements were used to infer the perceived location of the grating, with the tester tapping the corresponding screen quarter. Older children touched or pointed to "the smiley face" grating target (Figure 14A). Correct results produced a sound/animation reward involving a smiley face (Figure 14B). The tester held the tablet with the fingers positioned such that the test screen was not obscured (Figure 14C) In the absence of a

correct response, a lower- spatial frequency target was presented.

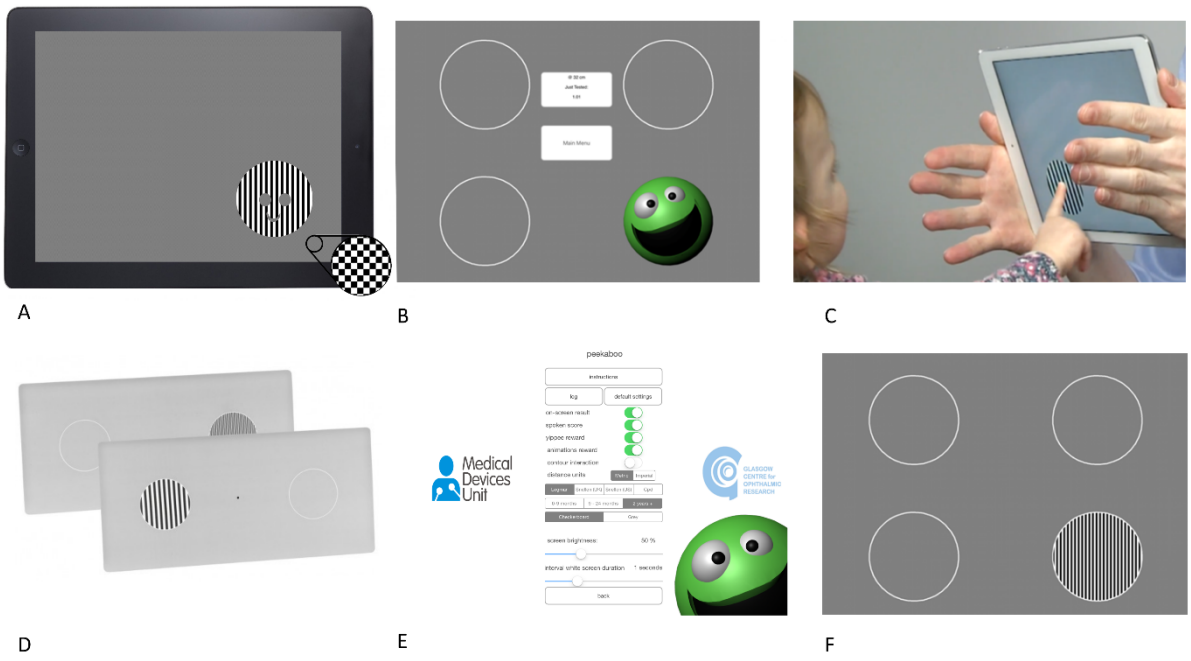


Figure 14. **A:** Peekaboo Vision build 1 (PVb1). **B**: Reward graphic and on-screen result. **C:** Demonstration of touchscreen functionality. **D:** The Keeler Acuity Test for Infants. **E:** Peekaboo Vision build for iOS (PVi) option configuration screen. **F:** Peekaboo Vision build for iOS (PVi) example test screen.

target spatial frequency (cpd)	acuity (logMAR equivalent)	Keeler cards @ 38 cm *additional set**	Peekaboo Vision (build 1) at 25 cm	Peekaboo Vision (iOS build) at:			
				25 cm	30 cm	40 cm	50 cm
51.0	-0.2						✓
35.4	-0.1	*				✓	
29.0	0.0				✓		
25.5	0.1	*	✓	✓			✓
21.6	0.2	*				✓	
14.5	0.3	*			✓		✓
12.5	0.4	✓	✓	✓		✓	✓
9.6	0.5	*			✓	✓	✓
7.7	0.6	*	✓	✓	✓	✓	✓
6.5	0.7	✓	✓	✓	✓	✓	✓
4.8	0.8		✓	✓	✓	✓	✓
3.8	0.9	*	✓	✓	✓	✓	✓
2.9	1.0	✓	✓	✓	✓	✓	✓
2.4	1.1		✓	✓	✓	✓	✓
2.1	1.2	*	✓	✓	✓	✓	✓
1.4	1.3	✓	✓	✓	✓	✓	✓
1.2	1.4		✓	✓	✓	✓	✓
1.0	1.5	*	✓	✓	✓	✓	✓
0.7	1.6	✓	✓	✓	✓	✓	✓
0.6	1.7		✓	✓	✓	✓	✓
0.5	1.8		✓	✓	✓	✓	✓
0.4	1.9	✓	✓	✓	✓	✓	✓
0.3	2.0	*	✓	✓	✓	✓	✓
0.2	2.2	✓	✓	✓	✓	✓	✓

Table 12. Spatial frequencies and equivalent acuity (logMAR) of gratings available with the Keeler Acuity Cards for Infants test plus additional cards, and with each of the builds of the Peekaboo Vision test.

The “additional set” relates to the more expansive set of cards, capturing more acuity levels than for more comprehensive testing. This was the set used in the present study. Cpd: cycles per degree.

5.2.1.4 Study protocol

Keeler Acuity Cards for Infants

Keeler Acuity Cards for Infants (KACI) were selected as the reference standard in this validation study, as it is the chosen test in our tertiary referral centre in Scotland and is used widely in UK practice. It is also the reference standard in other evaluations of digital grating paradigms [76]. While KACI does not have as extensive normative data as others, specifically Teller Acuity Cards [14,67], despite differences in design between both tests, Neu et al reported that Keeler Acuity Cards for Children and Teller Cards compared favourably on monocular testing in a slightly older, but similar age group to the present study (1-6 years old, N=95), with no significant differences in acuity scores[83].

An 18-card set was used, covering -0.1 to 2.0 logMAR (Figure 14D, range detailed in Table 12) including a blank card with no grating. The 1.5 logMAR card was used first, moving up/down the staircase depending on responses, as per instructions for use. For younger children, looking responses were judged by the tester.

Table 13 details the sizing of the grating elements for Keeler Acuity Cards for Infants and Peekaboo Vision.

	Keeler Acuity Cards for Infants		Peekaboo Vision		
	Physical size (mm)	Visual angle subtended (arcmin) at 38cm	Physical size (mm)	Visual angle subtended (arcmin) at:	
				25cm *	50cm (PVi) * †
Grating patch diameter	103	931	48	660	330
Grating patch center-to center	155	1402	H: 100 V: 76 D: 126	H: 1375 V: 1045 D: 1732	H: 688 V: 523 D: 866

Table 13. Physical properties/sizing details for Keeler Acuity Cards for Infants and Peekaboo Vision.

H: horizontal, V: vertical; D: diagonal.

*Indicates the distance used in both PVb1 and PVi builds;

†extended distance used at limit of dynamic range in PVi, Study 2.

All 58 children were tested with both Peekaboo Vision and Keeler cards. The both-eyes-open (BEO) condition was undertaken first, followed by right / left eye conditions (RE/LE), totalling six acuity tests. After a 15-minute interval, this process was repeated to assess test-retest repeatability. Each of the 58 children therefore underwent 12 acuity tests in total. Target distance was maintained using premeasured marks on the tester's arm. An engagement score of 0, 1 or 2 was awarded for every test result: 0 = no meaningful results; 1 = some meaningful results but loss of interest during test; and 2 = engagement to convincing threshold or finest grating.

5.2.2 Study 2

Methods for study 2 matched those for study 1, except for the aspects detailed below. Time-to-test was recorded for each test and viewing condition (BEO, RE, LE).

5.2.2.1 Patients

The West of Scotland Research Ethics Committee granted ethical approval. Sixty unselected children aged 2–60 months were recruited from the Royal Hospital for Sick Children eye clinic (Glasgow, UK), including those who exhibited visual problems as well as siblings free from visual problems.

5.2.2.2 Setting

Testing was performed in clinic rooms by an ophthalmologist (IL, SD) or senior orthoptist (AL, JWW).

5.2.2.3 Peekaboo Vision formal build for iOS (PVi)

A formal build used Swift 2 language for iOS versions 8.1 and above, scaled for the iPad3 platform. A video demonstration of this build is available [84]. The options screen is shown in Figure 14E. Changes from build 1 were informed by in-house testing and

study 1, comprising:

Randomisation (rather than pseudo-randomisation) of grating location

Automatic (rather than manual) overriding of brightness settings

Re-entry onto the rapid staircase if three consecutive levels are correct after an initial error (Figure 15B) to compensate for accidentally incorrect responses, sometimes caused by the tester inadvertently touching the screen

Removal of the smiley face details which created subtly visible edge artefacts within the grating (Figure 14A)

Addition of a ring around the grating target (Figure 14F), similar to Keeler cards, to limit visual cues where high spatial frequency gratings interact with the background

Configurable test distance, adjustable at cm increments within the range 25–50 cm (rather than fixed at 25 cm), to bypass screen resolution limitations, creating -0.2, -0.1, 0.0, 0.2 and 0.3 logMAR levels (Table 12)

Addition of a beep with each new target presentation to aid tester's recognition of synchronous looking responses coincident with appearance of grating

Addition of a feature to re-present the same target at a new random location in cases of equivocal responses, by shaking the device (akin to twirling a Keeler card).

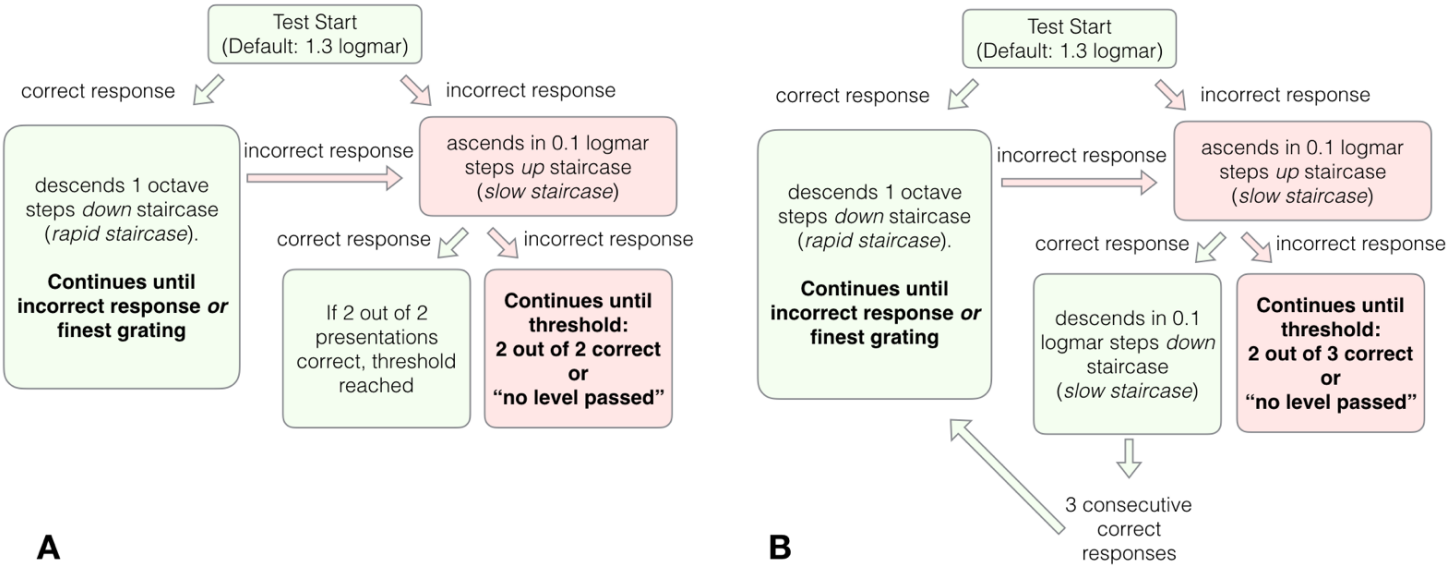


Figure 15. A: Stair-casing paradigm used in PVb1.

Correct responses occur when the child taps or points to the correct corner of the tablet, or the tester observes purposeful eye gaze and taps the corresponding corner of the tablet. Incorrect responses occur when child taps or points to the wrong corner, or the tester observes purposeful eye gaze and taps the incorrect corresponding corner, or there are no meaningfully directed eye movements. B: Stair-casing paradigm used in Peekaboo Vision build for iOS (PVi). The stair-casing change from the PVb1 version comprises automatic re-entry into the rapid staircase following 3 consecutive correct responses.

5.2.3 Statistics (Studies 1 and 2)

Engagement scores for the two test formats within each study (Keeler Cards versus PVb1 or PVi) were compared using McNemar's test for correlated proportions. Engagement scores (BEO versus monocular testing and also test 1 versus retest) were compared using Mann-Whitney U-tests. For test 1 versus retest comparison, BEO/RE/LE conditions were grouped together into test and retest groups.

Acuity scores were summarised using median values and compared between tests using limits of agreement and Bland-Altman plots. Test-retest repeatability was described using limits of agreement and coefficients of repeatability (CR; twice the standard deviation of the differences).

Test time was compared between PVi and Keeler Acuity Cards for Infants in study 2 using a paired t-test.

5.3 Results

5.3.1 Study 1

5.3.1.1 Engagement scores

Peekaboo Vision had higher engagement scores than Keeler cards over all three viewing conditions (BEO, RE, LE) (Table 14). An engagement score of 2 was achieved on significantly more occasions for Peekaboo Vision than for Keeler cards (100/174 (57%) versus 79/174 (45%), McNemar's test for correlated proportions, p=0.0005).

Engagement score with respect to age was analysed for BEO results for both tests. Thirty-six children attained an engagement score of 2 with Peekaboo Vision; their median age was 37 months. Thirty-one children attained an engagement score of 2 with Keeler cards; their median age was 38 months. Engagement scores of 1 were more frequent in younger children, with median ages of 28 months (N=23) and 25.5 months (N=18) for Keeler cards and Peekaboo Vision respectively. Engagement scores of 0 were infrequent, but in study 1 occurred in the same older children (N=4), with a median age of 36.5, using both Keeler cards and Peekaboo Vision.

There was evidence that engagement dropped slightly for Keeler card monocular testing following binocular testing: average engagement score dropped from median 2 (mean 1.5) for binocular testing (N=58) to median 1 (mean 1.3) for monocular testing (N=58, Mann-Whitney U-test, W=3665, adjusted p = 0.10). With Peekaboo Vision testing, engagement seemed to be marginally better maintained: average engagement score for binocular testing (N=58) dropped from median 2 (mean 1.6) to median 1.5 (mean 1.5) for monocular testing (N=58, Mann-Whitney U-test, W=3565, adjusted p = 0.3).

For Keeler cards, engagement score did not significantly drop between test and re-test, with a median score of 1 for both groups, with a modest decrease in the mean from 1.4 for test 1 versus 1.2 for retest (N=174, Mann-Whitney U-test, W=31977, adjusted

P=0.06).

For PVb1, ES did not significantly change between test and re-test, but evidenced a maintained median ES of 2 for across test and re-test, with the mean decreasing from 1.5 to 1.4 on re-test (N=174, Mann-Whitney U-test, W=31192, adjusted P=0.32).

		engagement score			p-value of difference in proportions of score=2
		0	1	2	
study 1 (N=58 subjects)	Keeler cards (N=174 measures)	16	79	79	0.0005
	Peekaboo Vision PVb1 (N=174 measures	11	63	100	
study 2 (N=60 subjects)	Keeler cards (N=158 measures)	7	28	123	0.5
	Peekaboo Vision PVi (N=158 measures	8	32	118	

Table 14. Engagement Scores

Number of tests (comprising BEO, RE and LE) with each engagement score (0, 1 or 2) for the two acuity tests. Figures are for first test, not retest. In the minority of instances where children had an abject loss of engagement, or became upset, no further testing was performed.

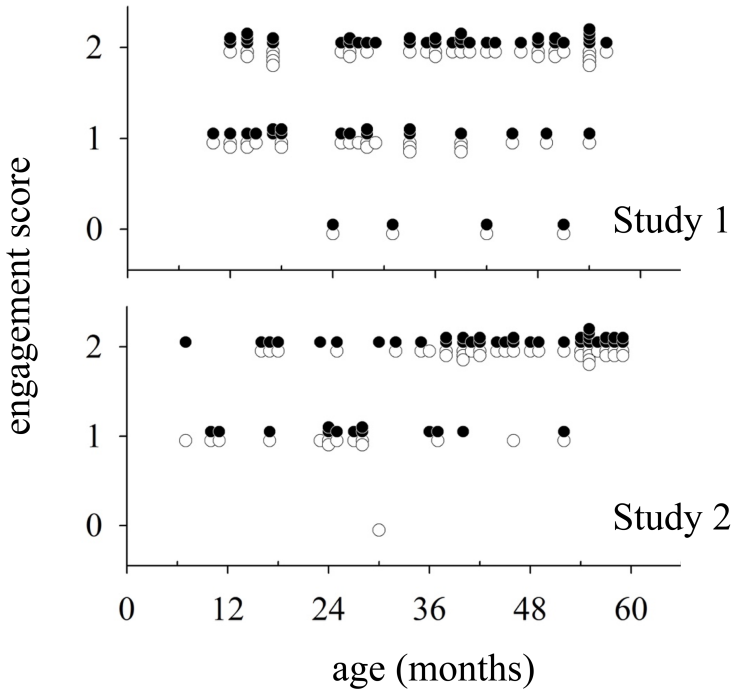


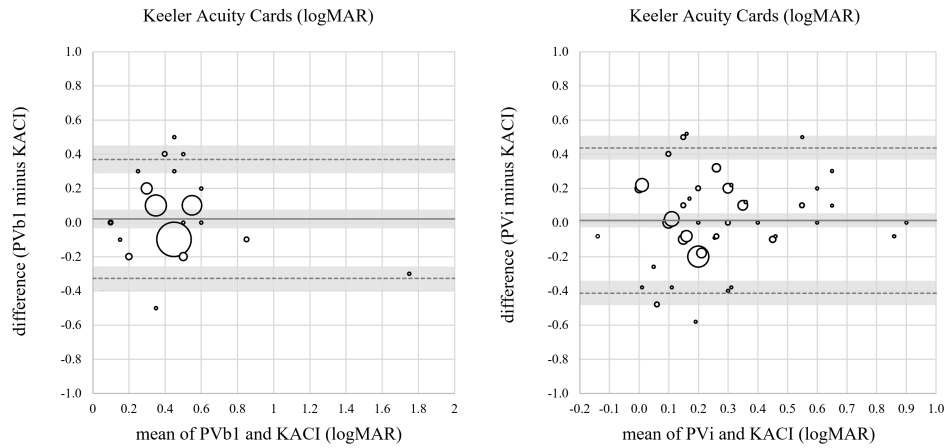
Figure 16. Scatterplots showing distribution of engagement scores with age of subjects for study 1 (upper panel) and study 2 (lower panel).

Data are for first tests, BEO viewing condition. Open circles: Keeler cards. Closed circles: Peekaboo Vision.

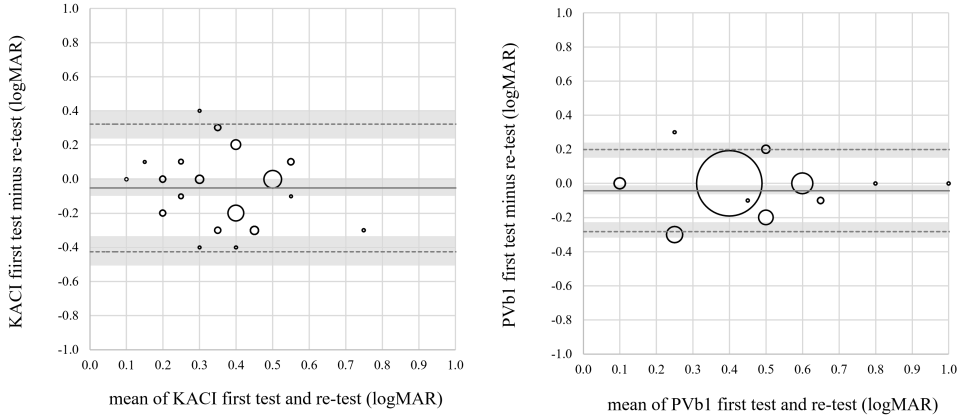
5.3.1.2 Acuity thresholds

For all viewing conditions (BEO, RE, LE), only children attaining an engagement score of 2 for both Peekaboo Vision and Keeler cards were included in comparative analysis. 72 tests from 31 children had engagement scores of 2 for both Peekaboo Vision and for Keeler cards. Median Peekaboo Vision acuity was 0.5 logMAR (range 0.1–1.9); median Keeler cards acuity was 0.4 logMAR (range 0.1–1.6). Agreement between the two tests was good, with a median absolute difference of 0.1 logMAR, mean difference of 0.02, 95%, LoA -0.33 to 0.37 LogMAR. The acuity difference between the two tests was ≤ 0.4 logMAR in 95% of tests and showed no tendency to vary with acuity (Figure 17A).

A



B



C

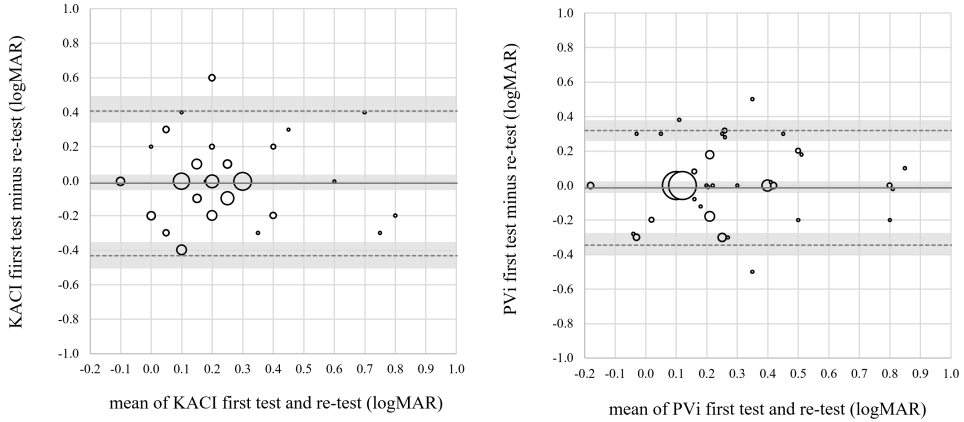


Figure 17. Bland Altman Plots for Studies 1 and 2

A: Bland-Altman plots of agreement. Left Panel: Study 1 (Malawi, Peekaboo Vision PVb1 versus KACI). Right Panel: Study 2 (UK, Peekaboo Vision PVi; right. Circles represent data points and are scaled to represent the number of instances the values occur. Solid horizontal lines represent mean difference, and dashed horizontal lines represent the limits of agreement. Shaded bands along mean difference and upper/lower limits of agreement illustrate 95% Confidence Intervals.

B: Test-retest repeatability in Study 1: Bland-Altman plots for Keeler cards (left) and for Peekaboo Vision (PVb1, right).

C: Test-retest repeatability in Study 2: Bland-Altman plots for Keeler cards (left) and for Peekaboo Vision (PVi, right).

5.3.1.3 Test-retest

To assess test-retest repeatability, pairs of tests with an engagement score of 2 at first test and again at re-testing some 15 minutes later, and for all viewing conditions (RE, LE, BEO) were compared. 85 pairs of Peekaboo Vision acuities and 58 pairs of Keeler card acuities were included (Figure 17B). Both tests showed good repeatability, with median differences between test and re-test being zero. Mean test-retest differences of -0.042 logMAR and -0.052 logMAR were found for Peekaboo Vision and for Keeler cards respectively. LoA were narrower and coefficients of repeatability (CR) were lower (better) for Peekaboo Vision (LoA -0.283 to 0.198 logMAR, CR 0.27) when compared to Keeler cards (LoA -0.427 to 0.323 logMAR, CR 0.37).

Repeatability was similar for Keeler binocular testing (N=20, LoA -0.359 to 0.399 logMAR, CR 0.39) and monocular testing (N=38, LoA -0.308 to 0.435 logMAR, CR 0.38). Repeatability was slightly poorer for Peekaboo binocular testing (N=31, LoA -0.270 to 0.315 logMAR, CR 0.30) than for monocular testing (N=54, LoA -0.148 to 0.244 logMAR, CR 0.20).

Test-retest acuity differences were compared for Keeler cards (N=20) and for Peekaboo vision (N=31) for BEO viewing conditions and engagement scores of 2: no significant differences were found, mean test-retest differences 0.020 vs 0.023 logMAR, 95%CI of difference -0.09–0.09, p=0.95.

5.3.2 Study 2

5.3.2.1 Engagement scores

Peekaboo Vision and Keeler cards had very similar engagement scores (Table 14), based on the first test result for all three viewing conditions (BEO, RE, LE). An engagement score of 2 was achieved for slightly fewer Peekaboo Vision tests than for Keeler card tests (118/158 (75%) versus 123/158 (77%)); McNemar's test for correlated proportions, p=0.5). Engagement scores for BEO results were reviewed with respect to age as for Study 1. Forty-seven children attained an engagement score of 2 with Peekaboo Vision; their median age was 54 months. Forty-five children attained an engagement score of 2 with Keeler cards; their median age was also 54 months. As for study 1, engagement scores of 1 were more frequent in younger children, with median ages of 27 months for both Peekaboo Vision (N=13) and Keeler cards (N=14). Engagement score of zero occurred only once, (17-month old) child using Keeler cards.

There was no convincing evidence that engagement dropped for monocular testing following binocular testing with Keeler cards: average engagement score for binocular testing (N=49) was median 2 (mean 1.8) and was median 2 (mean 1.7) for monocular testing (N=49, Mann-Whitney U-test, W=2505, adjusted p = 0.4). For Peekaboo Vision testing, engagement dropped slightly: average engagement score for binocular testing (N=49) was median 2 (mean 1.9) and was median 2 (mean 1.6) for monocular testing (N=49, Mann-Whitney U-test, W=2668, adjusted p = 0.02).

Regards change in engagement on test/retest, for Keeler cards, the median ES was maintained at 2, with the mean ES dropping modestly from 1.9 to 1.7 (N=119, Mann-Whitney U-test, W=14859, adjusted P=0.07).

Similarly, PVi did not show a significant change in ES on re-test, with median ES of 2

in both groups, mean decreasing from 1.9 to 1.8 on re-test (N=119, W=14763, adjusted P=0.10)

5.3.2.2 Time to test

The time-to-test data reflects the first test performed to limit any bias from learning or loss of interest from prolonged testing. Only those tests attaining engagement score of 2 (convincingly reached threshold without losing interest) with BEO were included. Mean time-to-test was just over a minute shorter for Peekaboo Vision than Keeler cards (N=33, 185 s vs 251 s, paired t-test p=0.002).

5.3.2.3 Acuity thresholds

For all viewing conditions (BEO, RE, LE), only those tests attaining an engagement score of 2 for both Peekaboo Vision and for Keeler cards were included in analysis. 110 tests from 46 infants and children had engagement scores of 2 for both Peekaboo Vision and Keeler cards. Median Peekaboo Vision acuity was 0.2 logMAR (range -0.18–0.90); median Keeler card acuity was 0.3 logMAR (range 0.10–0.90). Agreement between the two tests was good, with a median absolute difference of 0.18 logMAR (mean difference: 0.01, 95% LoA -0.413–0.437 logMAR). As would be expected in this population, the most frequently encountered acuities were in the normal range, between 0.0 and 0.4 logMAR, with no obvious bias across the range of acuities encountered (-0.18 – 0.9, Figure 17A, bottom right panel).

5.3.2.4 Test-retest

As for study 1, all pairs of tests with an engagement score of 2 at first test and again at re-testing, and for all viewing conditions (BEO, RE, LE) were assessed. 91 pairs of Peekaboo Vision acuities and 90 pairs of Keeler card acuities (Figure 17C) were compared. Both tests showed good repeatability, with median differences between test and re-test of zero and mean differences of -0.012 logMAR for both Peekaboo Vision and Keeler cards. As in study 1, LoA were narrower and CR were lower (better) for

Peekaboo Vision (LoA -0.344–0.320 logMAR, CR 0.32) than for Keeler cards (LoA -0.432–0.407 logMAR, CR 0.42).

Repeatability was similar for KACI binocular testing (N=36, LoA -0.429 to 0.429 logMAR, CR 0.44) and monocular testing (N=54, LoA -0.411 to 0.433 logMAR, CR 0.43). Repeatability was also similar for PVi binocular testing (N=33, LoA -0.262 to 0.383 logMAR, CR 0.33) and monocular testing (N=58, LoA -0.338 to 0.298 logMAR, CR 0.32).

Test-retest acuity differences were compared for KACI (N=36) and for PVi (N=33) for BEO viewing conditions and engagement scores of 2: no significant differences were found, mean test-retest differences 0.00 vs 0.06 logMAR, 95%CI of difference -0.15–0.03, p=0.2.

5.4 Discussion

In both studies, the mean difference in acuities measured by the card-based and by the digital test is modest, being 0.02 logMAR (95% CI for mean difference: -0.02 to 0.06) for Peekaboo Vision build 1 (study 1, Malawi), and 0.01 logMAR (95% CI for mean difference: -0.03 to 0.05) for Peekaboo Vision iOS build (study 2, UK). When comparing index with reference standard, the upper and lower limits of agreement (the interval of two standard deviations of the measurement differences either side of the mean difference) exceeded an octave step, but were within 5 logMAR lines (LoA 0.33-0.37 logMAR for Study 1; 95% LoA -0.413–0.437 logMAR for Study 2). These limits of agreement are similar to those observed when KACI is compared with itself on re-test in both studies (Study 1: LoA -0.427 to 0.323 logMAR; Study 2: LoA -0.432–0.407 logMAR).

Furthermore, with narrower limits of agreement on test-retest when compared with KACI in both studies, the findings support the use of high-resolution tablet-based technology, such as iPads, as a credible addition to the armamentarium available to clinicians in the assessment of vision in young children. However, as discussed below, there are limitations within these studies, and it cannot be unequivocally concluded that that these digital tests represent a replacement of the reference standard.

The number of forced-choice alternatives represents a major difference between the index tests and the 2-target Keeler Acuity Card. With PVb1, the staircase continued until 2-out-of-2 presentations were correct. This rigid staircasing did not allow progression when an error was made, and only poorer levels were tested thereafter. This method of staircasing reflects the recommended testing strategy for Keeler cards, as per instructions for use. Assuming a 1-in-4 (0.25) probability of the correct target being selected by pure chance at one presentation, the binomial probability of a given level being passed with 2 digital presentations by pure chance equals 0.063. Using a similar

paradigm, albeit with 2 targets per presentation (as is advised in the handbook accompanying Keeler Acuity Cards for Infants), the probability of 2 consecutive correct identifications at a given level arising by pure chance is 0.25.

With PVi, a change was made to the codebase altering the staircasing (Figure 15B): instead of terminating progression down the staircase after one error, and re-testing the previous level higher in the staircase, further presentations were given at the same level, and 2-out-of-3 correct test screen presentations (each with 4 potential positions) was required for a level to be passed.

Our expectation was that the more nuanced staircasing in PVi, allowing for staircase reversals, would increase accuracy. Given the significant differences in population and test builds between Studies 1 and 2, direct comparison of repeatability indices must be interpreted with caution, though it is interesting to note that PVb1 exhibited apparent superior indices of repeatability than PVi. This is likely to relate, in part, to the fewer testable levels at the finer range of acuities with PVb1, contributing to increased clustering around 0.1 and 0.4 levels (Figure 17B, left panel).

For PVi, the binomial probability of a level being passed with 2-out-of-3 correct identifications at a given spatial frequency grating, with random target selection, equals 0.141. If 2 consecutive presentations are correct for a given level (e.g. at the lower end of the staircase), a third presentation at that level is not offered. The probability of these 2 consecutive presentations being correctly identified by chance equals 0.063 (as with PVb1).

$$P(x) = \frac{N!}{x!\,(N-x)!}\pi^{x}(1-\pi)^{N-x}$$

P(x) = probability of x successes out of N trials, N = number of trials, π = probability of success on a given trial [85]

The 4-target setup in the index tests denotes an intrinsic lower probability of correct levels being passed by chance (0.063-0.141 for PVi and 0.063 PV) than Keeler Acuity Cards for Infants (0.25), which should theoretically increase reliability.

However, there are several other differences between the paradigms that are likely to impact on reliability. Having 4 target options on a smaller iPad screen makes re-fixation eye movement detection more challenging and demands a closer working distance to help mitigate for this. In turn, this brings the potential to positively bias towards myopes. Refractive status was not evaluated in either study, and would be a desirable aspect of future validation studies.

The closer working distance does bring another advantage, however, in that the screen can be reached by the child at arm's length. In this study, one recognised limitation in the methodology relates to the fact we did not record when children transitioned from a looking response to actively touching the gratings themselves. Although the examiner was not prescriptive regards looking/pointing/touching behaviour, the child was encouraged to tap the touchscreen. The "capture area" around the grating comprised one quarter of the total touchscreen, to help allow for minor target-touching "mis-hits" around a given grating. Touching the grating introduces another major divergence from

traditional card-based testing, where touching the gratings is discouraged due to the potential for permanently marking/scratching the card. Children were nevertheless encouraged to point to the KACI grating to limit bias. Given the age-groups involved, co-ordination of hand/arm movement presents another source of potential error, likely to be greater in younger infants and those with concurrent motor impairments.

The typical experience was that in older children, the natural tendency was to touch the grating. In less confident, typically younger children, the test frequently commenced as a preferential looking test, but after a few low- spatial frequency presentations, the behaviour often changed to pointing/touching. Recording such behaviour in future validation work would allow a more nuanced assessment of the accuracy of the digital platform specifically as a preferential looking test. Furthermore, attention to given behaviours within the context of varying acuity and age subgroups would help evaluate confounding influences.

For use as a preferential looking test, a potential shortcoming of the design of the PV test screens relates to the difference between vertical and horizontal spacing of the 4 test grating loci. This presents a potential bias whereby re-fixations horizontally/diagonally may be relatively easier to spot than vertical re-fixation eye movements. Indeed, this may exaggerate difficulties in deciding upon looking responses in cases of vertical or horizontal strabismus. Table 13 details the difference in visual angle between the digital platform and Keeler Cards. When tested at 25cm (compared to recommended 38cm test distance for Keeler Acuity Cards), the centre-to-centre distance of the digital grating patches is similar horizontally (1402 arcmin versus 1375 for PV), but reduced in the vertical plane to 1045 arcmin, around 25% less than that of the horizontal visual angle between the gratings in Keeler. These differences are more pronounced when the digital test distance is extended to the end of the dynamic range to 50cm, where the vertical visual angle reduces to just 523 arcmin, translating to a re-fixation eye movement

around 37% of the magnitude expected of keeler. While no examiners reported this to be an issue, a potential improvement in the design for future incarnations may be to match horizontal and vertical grating distances from centre, or limit the number of loci to 2 wider spaced target areas, which is an option within the iOS version of Peekaboo Vision, though not a setting that has been evaluated in the present study. Emerging large display platforms, such as the 12.9 inch Pro [86] provides over 77% increase in display area when compared to the model used in the present study, increasing the scope for incorporation of wider-spacing for targets for future versions of digital preferential looking tests.

Where looking responses are replaced with screen tapping or pointing by the child, the issues pertaining to detection of visual behaviour are largely obviated. Indeed, for such populations that can reliably tap proximally to the grating patch, there may be advantages to bringing the grating targets closer together, such that the gratings fall to a more central retinal position, providing a better index of central macular/foveal acuity.

Another noteworthy difference between the index and reference tests relate to the number of testable acuity levels in each (18 for Keeler and PV build 1 versus 25 for PV iOS build), as outlined in Table 12. This difference is also likely to impact on observed reliability differences. Furthermore, the step-size in the digital platform becomes more course towards the finer end of the high- spatial frequency grating range for the digital platforms. By clumping a large *range* of acuities between wide levels in one nominal acuity level, the precision would appear better than when compared with a test that captured more nuanced acuities *between* steps, but the accuracy may actually be poorer.

With the screen resolution and fixed test distance of 25cm in study 1, the highest two spatial frequencies possible are created with 1- and 2-pixel grating widths, corresponding with acuity scores of 0.12 and 0.42 logMAR (doubling of visual angle).

True acuities lying between 0.42 and 0.12 logMAR were therefore all scored as 0.42 logMAR (Table 1), resulting in the horizontal clustering of data at 0.42 logMAR in Figure 4B, right panel. In contrast, the Keeler cards with the Children's Additional Set included -0.1, 0.1, 0.2 and 0.3 logMAR: levels untestable with Peekaboo Vision build 1. This might be expected to have caused Peekaboo Vision to underestimate acuity in children with good acuities, thereby creating overall disagreement between Peekaboo Vision and Keeler cards, but this was not seen, perhaps because the numbers affected were small. For Peekaboo Vision iOS version (study 2), the adjustable test distance increased the range of measurable thresholds at the better acuity end of the test, but the differences noted between Peekaboo Vision and Keeler cards remained very small, suggesting there was little or no skew effect with Peekaboo build 1 in study 1.

In both studies, an animated smiley face comprised the reward animation, and during testing, after the animation reward was demonstrated, the connection between tapping the grating and the subsequent face animation was re-enforced to encourage the children to engage. Children with very poor vision, including those with central scotomas, may not have appreciated the details of the smiley face (within the grating in study 1, or within the animation reward in both studies), creating potential confusion. We did not observe any instances where this created an obvious barrier to testing, as even in those children with very poor vision, orientation of attention toward the lowest spatial frequency gratings on an otherwise featureless screen appeared instinctive. It should, however, be noted that statistical analysis regards acuity was confined to patients who were deemed reliable in reaching an endpoint with all tests (engagement score 2). In our data, this subset captured children with reasonable vision, with only one child exhibiting acuity poorer than 1.0 logmar. Consequently, impact of a child's inability to detect the face features (in PVb1 or reward in both Study 1 and 2) may be a factor missed in the present analysis. Future studies should evaluate PV in children with severe visual loss,

as it is not clear whether the present findings are generalizable.

Bittner et al [87] compared a digital gratings test (measuring up to 2.2 logMAR), which also employed a 4-target forced choice paradigm, comparing instead with ETDRS as gold standard in a population with low vision adults (legally blind) due to retinal disease. Interestingly, they demonstrated good agreement with digital gratings, which scaled similarly to ETDRS in their Retinitis Pigmentosa low vision group, with +/- 2SD within around +/- 0.4 logMAR steps.

When comparing index test (PVb1 and PVi) to reference standard, we observed around 0.4 logMAR difference in 95% of tests, which is similar to that observed when PVb1 was tested in an adult cohort with artificially degraded vision using +4, +8 or +16 spherical lenses to evaluate performance in low vision, as detailed in the preceding chapter [37].

Edge artefact at the junction between grating and high- spatial frequency background checkerboard presents a potential extra visual cue to children. This was a noticeable finding in PVb1 at the finest gratings, and also at the junction between grating and isoluminant background at the edges of the face elements (Figure 14A). A similar finding regards increased visibility in relation to grating edge effect was reported in relation to the Teller Acuity Test [88], which prompted the use of the white rings with Keeler Acuity Cards for Children[83]. Following this observation with PVb1 during Study 1, the white rings were included in PVi to remove this potential cue, and the face details used in PVb1 were removed.

The data presented here suggest Peekaboo Vision has better repeatability than Keeler preferential looking cards: coefficients of repeatability were 0.27 for Peekaboo Vision versus 0.37 for Keeler cards in study 1, and 0.32 for Peekaboo Vision versus 0.42 for Keeler cards in study 2. This is clinically important, particularly when measuring the

change in vision of a child over a period of time.

In study 1 (Malawi), children appeared to engage more with Peekaboo Vision than Keeler cards, while study 2 (UK) data suggested no difference. This may reflect changes made in the Peekaboo Vision application, such as the removal of the smiley face detail, or it may reflect differences in the populations tested: tablets are less widely available in Malawi than in the UK, so their relative novelty may have improved engagement. Other factors such as level of vision and ocular/medical conditions may have influenced engagement, but are difficult to quantify in such heterogeneous groups. Given that each child had to undertake up to 12 acuity tests (BEO, RE and LE test and retest for both Peekaboo Vision and Keeler cards) plus a 15-minute interval, better engagement might be expected for a single test in typical practice.

Our data suggests a trend towards decreased engagement on re-test across all comparisons, which is to be expected in the given age group. This did not however appear to reach statistical significance in our analysis in either study for either platform.

A drop in engagement score was also observed when testing monocularly after BEO testing, with mean engagement sore dropping from 1.9 to 1.6 (N=49, Mann-Whitney U-test, W=2668, adjusted p = 0.02). While statistically significant, the clinical significance of this small drop in engagement score is uncertain. It is possible that an intrinsic feature of PVi increases the potential for disinterest, introducing diminishing engagement with repeat testing. Future design adjustments could accelerate progression on the staircase and increase the variety of sounds and animation rewards to maintain engagement through binocular and monocular tests. Improvements in the study methodology are also desirable in future studies, limiting the testing gamut to replicate a test time more typical of a ‘real life’ clinical setting.

In Study 2, we note wider 95% limits of agreement for Keeler test/retest, approaching

an octave step, wider than that found in Study 1 with the same test. This may relate to several factors, particularly the very different populations, as well inter-tester variability. The extensive range of testing (12 tests) in such a young population is likely to be the most significant factor, and the most likely reason for the observed trend towards decreased ES across all repeat testing.

There are extensive differences between Study 1 and Study 2, both in population, as well as design of the index tests. Study 1 is pilot in nature, testing the methodology in advance of Study 2, and also informing the development of the formal Peekaboo Vision build for iOS. While it is useful to evaluate in broad terms how a digital infant acuity test performs in Malawi, we cannot draw any meaningful conclusions based on comparisons between the 2 studies due to intrinsic differences between PVb1 and PVi. The next logical study would seek to repeat the methodology with PVi in a similar cohort in Malawi.

Only total session time was noted for study 1. For study 2, the data inclusion form was amended to capture individual test times. The mean time-to-test (first test, BEO) was over 1 minute shorter for Peekaboo Vision than for Keeler cards (185 s vs 251 s, paired t test, N=33, p= 0.0021), i.e. 26% shorter. This may be partly due to Peekaboo Vision's four-choice paradigm compared with Keeler cards' two-choice paradigm. Shorter test-time represents an important potential benefit of Peekaboo Vision given the short attention span of this age-group; another is the potential cost saving in high throughput orthoptic clinics or screening programs.

Compared to card-based vision tests, tablet-based tests are susceptible to veiling glare and to reflections, preventing use outside. On the other hand, a tablet's Lambertian surface can maintain contrast even if viewing angle is not perpendicular, and photometric compliance of tablets with British and European Standards is at least as

good as gold standard retro-illuminated ETDRS charts, as discussed in Chapter 2 [63], with the potential advantages detailed in Chapter 3. Given the potential for variation between reflected light from cards and Lambertian displays, measuring ambient luminance, together with reflected/emitted light from the cards/iPads respectively would have been useful area to explore, but was not addressed in the present research, which aimed to evaluate precision and accuracy in a "real life" context. The impact of varying lighting conditions was evaluated in Chapter 3 and appear not to be an issue regards luminance and contrast for either digital or traditional platforms.

5.5 Conclusion

Ongoing regulatory checks of applications for such measures are desirable given the frequent updates to operating systems and hardware. Expansion of National and International standards for vision testing equipment to include such ubiquitous mobile technology could help support the safe adoption of tablet-based vision testing into regular practice. Further investigation is required to evaluate the role of the technology in amblyopia screening, and evaluate performance in non-expert testers.

Chapter 6: Conclusions

The present research has systematically evaluated the utility of the computer tablet as a means to test vision in an infant population.

6.1 Chapter 1

Chapter 1 has outlined the scope of the problem amblyopia presents from a global perspective and overviewed the literature around the importance of early intervention.

The present reference standard infant acuity testing have been introduced, and literature around the disparity between grating acuity norms for card-based tests have between tests reviewed. With the present card-based standards defined, emerging mobile technology-based platforms were introduced.

6.2 Chapter 2

Chapter 2 evaluated the physical properties of a leading computer tablet display in relation to National and International standards for letter acuity chart design. The results from this study demonstrated that the photometric standards of luminance, contrast, and luminance uniformity are met more effectively by the range of iPads under test than by dedicated ETDRS charts in active clinical use in a tertiary referral unit. The study met its aim of documenting the suitability of this mobile technology regards high contrast acuity standards, providing a practical guide to health care professionals working within eye care.

6.3 Chapter 3

Chapter 3 extended a similar analysis to the field of infant acuity chart based testing, where the regulatory standards that pertain to adult letter acuity charts are, at the time of writing, absent.

Given the intrinsic pivot within infant based card testing is the relative visibility of high contrast targets on an isoluminant background, together with the greater consequences

of failed detection of a visual problem in a paediatrics population, it could be argued that the need for standardisation in this area exceeds that of adult acuity platforms. The studies within Chapter 3 unpack the differences across available platforms and add digital platforms to the range tested, allowing a comparative analysis. The findings demonstrate intrinsic advantages to digital platforms regards achieving a mean average luminance "grey" background relative to the black and white values of composite foreground gratings. The experiments highlight marked heterogeneity across traditional card-based platforms regards the handling of the grey backdrop, which we theorise is an important factor contributing to the wide variation across platforms [30–33].

6.4 Chapter 4

With the context set regards physical properties of ETDRS charts, digital screens and infant acuity cards, and how they compare with present chart-design standards, Chapter 4 brings together observations around digital handling of black, white and grey background elements into a prototype digital acuity test, Peekaboo Vision Build 1 (PVb1), and evaluates the psychophysiological context, demonstrating proof of concept regards a grating acuity test being delivered by a digital tablet platform in an adult cohort. With blurring conditions, the prototype application, PVb1 performed well across a range of artificially degraded acuities, with observed potential benefits regards test-retest repeatability.

6.5 Chapter 5

Chapter 5 combines 2 paediatric test comparison trials. Study 1 translated the prototype build described in Chapter 4, PVb1, to a remote African setting, and evaluated performance agains KACI as the reference standard. Other indices measured included the engagement of the child. Study 2 comprised a similar methodology to Study 1 but had important differences in terms of test enhancements (informed by experience with

Study 1), transitioning from PVB1 to Peekaboo Vision for iOS (PVi). Extending the evaluation to a Western population of similarly aged young children.

Across both studies, the mean difference between reference standard and digital version was modest (-0.03 to 0.01), with notable differences in upper and lower limits of agreement in favour of the digital platform (exhibiting narrower LoA). Peekaboo Vision evidenced improved repeatability than Keeler preferential looking cards: coefficients of repeatability were 0.27 for Peekaboo Vision versus 0.37 for Keeler cards in study 1, and 0.32 for Peekaboo Vision versus 0.42 for Keeler cards in study 2. This is clinically important, particularly when measuring the change in vision of a child over a period of time. The mean time-to-test was over 1 minute shorter (26%) for Peekaboo Vision than for Keeler cards (p= 0.0021) Shorter test-time represents an important potential benefit of Peekaboo Vision given the short attention span of this age-group; another is the potential cost saving in high throughput orthoptic clinics or screening programs.

The source of these differences reported in Chapter 5, however, cannot solely be attributed to the physical properties of the tested platforms, as observed differences are likely to arise from a multifaceted divergence in test designs: sound and animation rewards in the digital platforms, differing test distances, differing layouts, differing stair casing paradigms, differing numbers of presented optotypes at one time, etc. Nevertheless, taken together, the results demonstrate utility in both high income and rural low/middle settings where traditional equipment cannot be financed.

6.6 Future Work

Integration of eye-tracking on a monitor-based system has also been evaluated with grating acuity in children[76], and could increase objectivity and potentially remove the need for a stringent fixed test distance, with live distance measuring between eye and

tracker, dynamically adjusting acuity score relative to distance at the moment of re-fixation.

Furthermore, using a large crowd-sourced data set to train a deep learning convolutional network, the native front-facing mobile camera alone has been demonstrated to predict gaze with an accuracy purported to outperform current state of the art approaches [89].

Combining such technology with the high fidelity lambertion tablet display, and using gaze to guide preferential looking methodologies through an automated staircase culminates in an exciting possibility, whereby visual function could be profiled using a software-only solution on a near ubiquitous mobile platform designed for recreation. Such development could extend the reach of a visual screening program into the patient's home. Indeed, such directions present an exciting new direction for paediatric vision testing, not just for high-contrast acuity, but for contrast and colour assessment [38,40].

Future work within this area will seek to mount infant vision tests within emerging Virtual Reality Headsets. Advantages include a fixed test distance, controlled microenvironment, wide field of view, leveraging of inertial sensors towards tracking of pursuit of target within detection thresholds. At the time of writing, the bottleneck to mounting such a test relates to the relative low resolution of available devices, though this is undoubtedly set to become a solved problem as the rapid pace of digital display and pico-projector development technology rapidly evolves.

Regulatory guidance in this area would go some way to harmonise practice and provide a framework around a minimum standard, allowing the high standards that exist for adult testing (ETDRS) to be extended to the infant population, and to evolving digital platforms.

References

1. Gilbert C, Foster A. Childhood blindness in the context of VISION 2020--the right to sight. Bull World Health Organ. 2001;79(3):227–32.

2. Maurer D, Lewis TL, Brent HP, Levin AV. Rapid Improvement in the Acuity of Infants After Visual Input. Science. 1999 Jan 10;286(5437):108–10.

3. Daw NW. Critical periods and amblyopia. Arch Ophthalmol. 1998 Apr 1;116(4):502–5.

4. Wallace DK, Lazar EL, Melia M, Birch EE, Holmes JM, Hopkins KB, et al. Stereoacuity in children with anisometropic amblyopia. Journal of American Association for Pediatric Ophthalmology and Strabismus. 2011 Oct 1;15(5):455–61.

5. Holmes JM, Clarke MP. Amblyopia. The Lancet. 2006 Apr 22;367(9519):1343–51.

6. Rahi JS, Logan S, Timms C, Russell-Eggitt I, Taylor D. Risk, causes, and outcomes of visual impairment after loss of vision in the non-amblyopic eye: a population-based study. The Lancet. 2002 Aug 24;360(9333):597–602.

7. Opthalmologists TRC of. Children's Eye Care – Reinforcing National Guidelines for Vision Screening in Schools for all 4-5 year olds [Internet]. The Royal College of Ophthalmologists. 2015 [cited 2019 Jun 27]. Available from: https://www.rcophth.ac.uk/2015/09/childrens-eye-care-reinforcing-national-guidelines-for-vision-screening-in-schools-for-all-4-5-year-olds/

8. One in five PCTs do not offer school sight tests [Internet]. Which? News. 2011 [cited 2019 Jun 27]. Available from: https://www.which.co.uk/news/2011/08/one-in-five-pcts-do-not-offer-school-sight-tests-264291/

9. Taylor K, Elliott S. Interventions for strabismic amblyopia. Cochrane Database of Systematic Reviews [Internet]. 2014 [cited 2019 Jun 9];(7). Available from: https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD006461.pub4/full

10. Flynn JT, Woodruff G, Thompson JR, Hiscox F, Feuer W, Schiffman J, et al. The therapy of amblyopia: an analysis comparing the results of amblyopia therapy utilizing two pooled data sets. Trans Am Ophthalmol Soc. 1999;97:373–95.

11. Williams C, Northstone K, Harrad RA, Sparrow JM, Harvey I. Amblyopia treatment outcomes after screening before or at age 3 years: follow up from randomised trial. BMJ. 2002 Jun 29;324(7353):1549.

12. Espinosa JS, Stryker MP. Development and Plasticity of the Primary Visual Cortex. Neuron. 2012 Jul 26;75(2):230–49.

13. Teller DY. The forced-choice preferential looking procedure: A psychophysical technique for use with human infants. Infant Behav Dev. 1979 Jan 1;2:135–53.

14. Teller DY, McDonald MA, Preston K, Sebris SL, Dobson V. Assessment of Visual Acuity in Infants and Children: the Acuity Card Procedure. Dev Med Child Neurol. 1986 Dec;28(6):779–89.

15. Adoh TO, Woodhouse JM, Oduwaiye KA. The Cardiff Test: a new visual acuity test for toddlers and children with intellectual impairment. A preliminary report. Optom Vis Sci. 1992 Jun;69(6):427–32.

16. Dobson V, McDonald MA, Teller DY. Visual Acuity of Infants and Young Children: Forced-Choice Preferential Looking Procedures. Am Orth J. 1985 Jan 1;35(1):118–25.

17. Brown JM, Campbell JP, Beers A, Chang K, Ostmo S, Chan RVP, et al. Automated Diagnosis of Plus Disease in Retinopathy of Prematurity Using Deep Convolutional Neural Networks. JAMA Ophthalmol. 2018 Jul 1;136(7):803–10.

18. Fauw JD, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat Med. 2018 Sep;24(9):1342–50.

19. Lea Gratings test for toddlers (Art. No. 81047) [Internet]. VISUS Sehtestprodukte. [cited 2019 Jun 23]. Available from: https://www.visus-sehteste.de/product_info.php?info=p372_lea-gratings-test-for-toddlers--art--no--81047-.html

20. Keeler Acuity Cards Technical Bulletin [Internet]. [cited 2015 Dec 22]. Available from: http://www.richmondproducts.com/files/4913/7832/4419/KAC_Childrens_Grating_Test_Tech_Bulletin_no_price_082713.pdf

21. Marketing C. Testing and Administrating with Teller Acuity Cards II [Internet]. 2014 [cited 2019 Jun 23]. Available from: https://vimeo.com/101310031

22. TELLER ACUITY SET, 16 plates Eyesfirst.eu [Internet]. [cited 2019 Jun 23]. Available from: https://www.eyesfirst.eu/teller-acuity-cards.html?___store=en

23. Cardiff Acuity Test flyer 102505.pdf [Internet]. [cited 2019 Jun 23]. Available from: http://media.ophthalmologyweb.com/m/27/Article/productliterature/Cardiff%20Acuity%20Test%20flyer%20102505.pdf

24. Howland B, Ginsburg A, Campbell F. High-pass spatial frequency letters as clinical optotypes. Vision Res. 1978;18(8):1063–6.

25. Zanker J, Mohn G, Weber U, Zeitler-Driess K, Fahle M. The development of vernier acuity in human infants. Vision Research. 1992 Aug 1;32(8):1557–64.

26. Norcia AM, Tyler CW, Hamer RD. Development of contrast sensitivity in the human infant. Vision Research. 1990 Jan 1;30(10):1475–86.

27. Fawcett SL, Wang Y-Z, Birch EE. The Critical Period for Susceptibility of Human Stereopsis. Invest Ophthalmol Vis Sci. 2005 Feb 1;46(2):521–5.

28. Simmers AJ, Ledgeway T, Hess RF, McGraw PV. Deficits to global motion processing in human amblyopia. Vision Research. 2003 Mar 1;43(6):729–38.

29. Adams RJ, Courage ML, Mercer ME. Systematic measurement of human neonatal color vision. Vision Res. 1994 Jul;34(13):1691–701.

30. Sharma P, Bairagi D, Sachdeva MM, Kaur K, Khokhar S, Saxena R. Comparative evaluation of teller and cardiff acuity tests in normals and unilateral amblyopes in under-two-year-olds. Indian J of Ophthalmol. 2003 Jan 12;51(4):341.

31. Kushner BJ, Lucchese NJ, Morton GV. Grating Visual Acuity With Teller Cards Compared With Snellen Visual Acuity in Literate Patients. Arch Ophthalmol. 1995 Apr 1;113(4):485–93.

32. Mody KH, Kothari MT, Sil A, Doshi P, Walinjkar JA, Chatterjee D. Comparison of lea gratings with cardiff acuity cards for vision testing of preverbal children. Indian J Ophthalmol. 2012;60(6):541–3.

33. Shah N, Dakin SC, Redmond T, Anderson RS. Vanishing Optotype acuity: repeatability and effect of the number of alternatives. Ophthalmic and Physiological Optics. 2011;31(1):17–22.

34. Mobile Health Trends and Figures 2013-2017 [Internet]. research2guidance. [cited 2019 Jun 22]. Available from: https://research2guidance.com/product/mobile-health-trends-and-figures-2013-2017/

35. WHO | Universal eye health: a global action plan 2014–2019 [Internet]. WHO. [cited 2019 Jun 23]. Available from: http://www.who.int/blindness/actionplan/en/

36. Bastawrous A, Rono HK, Livingstone IAT, et al. Development and validation of a smartphone-based visual acuity test (peek acuity) for clinical practice and community-based fieldwork. JAMA Ophthalmol [Internet]. 2015 May 28 [cited 2015 Jun 4]; Available from: http://dx.doi.org/10.1001/jamaophthalmol.2015.1468

37. Livingstone IAT, Lok ASL, Tarbert C. New mobile technologies and visual acuity. Conf Proc IEEE Eng Med Biol Soc. 2014;2014:2189–92.

38. Aslam TM, Murray IJ, Lai MYT, Linton E, Tahir HJ, Parry NRA. An assessment of a modern touch-screen tablet computer with reference to core physical characteristics necessary for clinical vision testing. J R Soc Interface. 2013 Jul 6;10(84):20130239.

39. Aslam TM, Parry NRA, Murray IJ, Salleh M, Col CD, Mirza N, et al. Development and testing of an automated computer tablet-based method for self-testing of high and low contrast near visual acuity in ophthalmic patients. Graefes Arch Clin Exp Ophthalmol. 2016 Feb 22;1–9.

40. Tahir HJ, Murray IJ, Parry NRA, Aslam TM. Optimisation and Assessment of Three Modern Touch Screen Tablet Computers for Clinical Vision Testing. PLoS ONE. 2014 Apr 23;9(4):e95074.

41. Black JM, Jacobs RJ, Phillips G, Chen L, Tan E, Tran A, et al. An assessment of the iPad as a testing platform for distance visual acuity in adults. BMJ Open. 2013;3(6).

42. Bastawrous A, Cheeseman RC, Kumar A. iPhones for eye surgeons. Eye (Lond). 2012 Mar;26(3):343–54.

43. Medical devices: software applications (apps) [Internet]. GOV.UK. [cited 2019 Jun 23]. Available from: https://www.gov.uk/government/publications/medical-devices-software-applications-apps

44. Personalised health and care 2020: a framework for action [Internet]. GOV.UK. [cited 2019 Jun 23]. Available from: https://www.gov.uk/government/publications/personalised-health-and-care-2020/using-data-and-technology-to-transform-outcomes-for-patients-and-citizens

45. McDonald MA, Dobson V, Sebris SL, Baitch L, Varner D, Teller DY. The acuity card procedure: a rapid test of infant acuity. Invest Ophthalmol Vis Sci. 1985 Aug 1;26(8):1158–62.

46. Ruamviboonsuk P, Sudsakorn N, Somkijrungroj T, Engkagul C, Tiensuwan M. Reliability of visual acuity measurements taken with a notebook and a tablet computer in participants who were illiterate to Roman characters. J Med Assoc Thai. 2012 Mar;95 Suppl 3:S109-116.

47. Zhang Z, Zhang S, Huang X, Liang L. A pilot trial of the iPad tablet computer as a portable device for visual acuity testing. J Telemed Telecare. 2013 Jan;19(1):55–9.

48. Sheedy JE, Bailey IL, Raasch TW. Visual acuity and chart luminance. Am J Optom Physiol Opt. 1984 Sep;61(9):595–600.

49. Johnson CA, Casson EJ. Effects of luminance, contrast, and blur on visual acuity. Optom Vis Sci. 1995 Dec;72(12):864–9.

50. International Council of Ophthalmology : Resources : Visual Acuity Measurement Standard [Internet]. [cited 2019 Jun 23]. Available from: http://www.icoph.org/resources/47/Visual-Acuity-Measurement-Standard.html

51. BS 4274-1:2003 - Visual acuity test types. Test charts for clinical determination of distance visual acuity. Specification [Internet]. [cited 2019 Jun 23]. Available from: https://shop.bsigroup.com/ProductDetail/?pid=000000000030021439

52. Information Display Measurements Standard 1.03 | 3D-ConTourNet [Internet]. [cited 2019 Jun 23]. Available from: https://www.3d-contournet.eu/information-display-measurements-standard-1-03/

53. Samei E, Badano A, Chakraborty D, Compton K, Cornelius C, Corrigan K, et al. Assessment of display performance for medical imaging systems: Executive summary of AAPM TG18 report: Performance assessment of medical displays. Med Phys. 2005 Apr;32(4):1205–25.

54. International Council of Ophthalmology : Resources : Visual Acuity Measurement Standard [Internet]. [cited 2019 Jun 23]. Available from: http://www.icoph.org/resources/47/Visual-Acuity-Measurement-Standard.html

55. Apple - iPad Air 2 - Design [Internet]. Apple. [cited 2015 Jun 15]. Available from: http://www.apple.com/ipad-air-2/design/

56. Shimpi AL. iPad 4 (Late 2012) Review [Internet]. [cited 2019 Nov 4]. Available from: https://www.anandtech.com/show/6472/ipad-4-late-2012-review

57. B H, H O, E Z. ["Glare vision". I. Physiological principles of vision change with increased test field luminance]. Klin Monbl Augenheilkd. 1992 Feb;200(2):105–9.

58. Bühren J, Terzi E, Bach M, Wesemann W, Kohnen T. Measuring contrast sensitivity under different lighting conditions: comparison of three tests. Optom Vis Sci. 2006 May;83(5):290–8.

59. Replacement lamp for 2425 purchased before Nov 2008 [Internet]. [cited 2015 Jun 15]. Available from: http://precision-vision.com/all-products/illuminator-cabinets/etdrs-illuminator-cabinet-and-accessories/replacement-lamp-for-cat-no-2425.html

60. OpenCV [Internet]. [cited 2019 Jun 23]. Available from: https://opencv.org/

61. Bastawrous A, Giardini ME, Bolster NM, Peto T, Shah N, Livingstone IAT, et al. Clinical Validation of a Smartphone-Based Adapter for Optic Disc Imaging in Kenya. JAMA Ophthalmol. 2016 Feb;134(2):151–8.

62. Bodduluri L, Boon MY, Ryan M, Dain SJ. Normative values for a tablet computer-based application to assess chromatic contrast sensitivity. Behav Res. 2018 Apr 1;50(2):673–83.

63. Livingstone IAT, Tarbert CM, Giardini ME, Bastawrous A, Middleton D, Hamilton R. Photometric Compliance of Tablet Screens and Retro-Illuminated Acuity Charts As Visual Acuity Measurement Devices. PLOS ONE. 2016 Mar 22;11(3):e0150676.

64. App Screenshot [Internet]. [cited 2017 Jun 13]. Available from: https://lh3.googleusercontent.com/R9ZA9qAOSR4z-CSYCjcpjrH_yuw6AQBWzE7zgD9iLOYKK74sW4npxWlbXyrPD_zAqZA=h310-rw

65. Atkinson J, Braddick O, Pimm-Smith E. 'Preferential looking' for monocular and binocular acuity testing of infants. Brit J of Ophthalmol. 1982 Apr 1;66(4):264–8.

66. Braddick O, Atkinson J. Development of human visual function. Vision Research. 2011 Jul 1;51(13):1588–609.

67. Mayer DL, Beiser AS, Warner AF, Pratt EM, Raye KN, Lang JM. Monocular acuity norms for the Teller Acuity Cards between ages one month and four years. Invest Ophthalmol Vis Sci. 1995 Mar;36(3):671–85.

68. Elgohary AA, Abuelela MH, Eldin AA. Age norms for grating acuity and contrast sensitivity measured by Lea tests in the first three years of life. Int J Ophthalmol. 2017 Jul 18;10(7):1150–3.

69. Martini G, Netto AA, Morcillo AM, Gagliardo HGRG, de Oliveira DF. The LEA Grating Test in assessing detection grating acuity in normal infants less than 4 months of age. J AAPOS. 2014 Dec;18(6):563–6.

70. Goren CC, Sarty M, Wu PY. Visual Following and Pattern Discrimination of Face-like Stimuli by Newborn Infants. Pediatrics. 1975 Oct;56(4):544–9.

71. Tanaka JW, Farah MJ. Parts and wholes in face recognition. Q J Exp Psychol A. 1993 May;46(2):225–45.

72. The Technology That's Making a Difference in the Developing World [Internet]. USGLC. [cited 2019 Jun 27]. Available from: https://www.usglc.org/blog/the-technology-thats-making-a-difference-in-the-developing-world/

73. Bolster NM, Giardini ME, Livingstone IA, Bastawrous A. How the smartphone is driving the eye-health imaging revolution. Exp Rev Ophthalmol. 2014 Dec 1;9(6):475–85.

74. Dorr M, Lesmes LA, Lu Z-L, Bex PJ. Rapid and reliable assessment of the contrast sensitivity function on an iPad. Invest Ophthalmol Vis Sci. 2013 Nov 5;54(12):7266–73.

75. Hathibelagal AR, Leat SJ, Irving EL, Nandakumar K, Eizenman M. Measuring Infant Visual Acuity with Gaze Tracker Monitored Visual Fixation. Optom Vis Sci. 2015 Jul;92(7):823–33.

76. Jones PR, Kalwarowsky S, Atkinson J, Braddick OJ, Nardini M. Automated Measurement of Resolution Acuity in Infants Using Remote Eye-Tracking. Investigative Ophthalmol Vis Sci. 2014 Dec 12;55(12):8102–10.

77. Kollbaum PS, Jansen ME, Kollbaum EJ, Bullimore MA. Validation of an iPad test of letter contrast sensitivity. Optom Vis Sci. 2014 Mar;91(3):291–6.

78. Wroblewski D, Francis BA, Sadun A, Vakili G, Chopra V. Testing of Visual Field with Virtual Reality Goggles in Manual and Visual Grasp Modes. BioMed Research International. 2014 Jun 23;2014:e206082.

79. The Stroke Vision App: A Screening Tool for Visual Stroke - Full Text View - ClinicalTrials.gov [Internet]. [cited 2017 Jul 25]. Available from: https://clinicaltrials.gov/ct2/show/NCT02539381

80. Murray I, Perperidis A, Brash H, Cameron L, McTrusty A, Fleck B, et al. Saccadic Vector Optokinetic Perimetry (SVOP): a novel technique for automated static perimetry in children using eye tracking. Conf Proc IEEE Eng Med Biol Soc. 2013;2013:3186–9.

81. Eye Tracking for Everyone.pdf [Internet]. Google Docs. [cited 2017 Nov 21]. Available from: https://drive.google.com/file/d/0B9HHfYresOgRaThhbEx1ek9POWM/view?usp=embed_facebook

82. Occluding Glasses For Pediatric Vision Testing [Internet]. Kay Pictures. [cited 2018 May 10]. Available from: http://kaypictures.co.uk/product/occluding-glasses/

83. Neu B, Sireteanu R. Monocular acuity in preschool children: Assessment with the Teller and Keeler acuity cards in comparison to the C-test. Strabismus. 1997;5(4):185–202.

84. iain livingstone. Peekaboo Vision for iPad (UK) [Internet]. Available from: https://www.youtube.com/watch?v=---Ow3lgj2w

85. Binomial Distribution [Internet]. [cited 2018 May 8]. Available from: http://onlinestatbook.com/2/probability/binomial.html

86. iPad Pro - Technical Specifications [Internet]. Apple (United Kingdom). [cited 2018 May 9]. Available from: https://www.apple.com/uk/ipad-pro/specs/

87. Bittner AK, Jeter P, Dagnelie G. Grating Acuity and Contrast Tests for Clinical Trials of Severe Vision Loss. Optom Vis Sci. 2011 Oct;88(10):1153–63.

88. Robinson J, Moseley MJ, Fielder AR. Grating acuity cards: spurious resolution and the 'edge artifact'. Clin Vision Sci. 1998;3:285–8.

89. Krafka K, Khosla A, Kellnhofer P, Kannan H, Bhandarkar S, Matusik W, et al. Eye Tracking for Everyone. In 2016 [cited 2017 Nov 21]. p. 2176–84. Available from: https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Krafka_Eye_Tracking_for_CVPR_2016_paper.html

Appendix A: IEEE Correspondence

From: customercare@copyright.com
Subject: Case #00781482 - Re-use of paper in entirety in MD thesis [ref:_00D30oeGz._5000c1pws6C:ref]
Date: 31 May 2019 at 05:41
To: iain.livingstone@nhs.net

Dear Iain Livingstone,

Thank you for contacting Copyright Clearance Center's Rightslink service. RightsLink provides copyright permission for the reproduction and distribution of copyrighted materials in print and electronic formats on behalf of copyright owners who list their titles with us.

I checked and found this copyright statement:

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

Please follow above instructions and guide lines for your permission request.

If you have any further questions please don't hesitate to contact a Customer Account Specialist at 855-239-3415 Monday-Friday, 24 hours/day.

Best,
Nadine Sobusa
Customer Account Specialist
Copyright Clearance Center
222 Rosewood Drive
Danvers, MA 01923
www.copyright.com
Toll Free US +1.855.239.3415
International +1.978-646-2600
Facebook - Twitter - LinkedIn

ref:_00D30oeGz._5000c1pws6C:ref

---------- Original Message ----------
From: LIVINGSTONE, Iain [iain.livingstone@nhs.net]
Sent: 5/30/2019 5:16 PM
To: customercare@copyright.com
Subject: Re-use of paper in entirety in MD thesis

Dear Customer Services Team,

I am first author on 2014 paper: New mobile technologies and visual acuity, published in 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society Conference Proceedings.

My reading of the helpful advice on [https://s100.copyright.com/AppDispatchServlet#formTop] is that with the stipulated

Appendix B: TVST Correspondence

From: **Debbie Chin** dchin@arvo.org
Subject: RE: Permission to use published content within MD thesis
Date: 31 May 2019 at 18:12
To: LIVINGSTONE, Iain (NHS FORTH VALLEY) iain.livingstone@nhs.net

DC

Dear Iain Livingstone,

Thank you for your email. You may reuse your article in your thesis as described below, with clear attribution.

Best regards,

Debbie Chin
ARVO Journals
1801 Rockville Pike, Suite 400
Rockville, MD 20852
+1.240.221.2926

From: LIVINGSTONE, Iain (NHS FORTH VALLEY) <iain.livingstone@nhs.net>
Sent: Thursday, May 30, 2019 6:08 PM
To: Debbie Chin <dchin@arvo.org>
Subject: Permission to use published content within MD thesis

Dear TVST Team

Many thanks for your help.

In January this year, we had our article "Testing Pediatric Acuity With an iPad: Validation of “Peekaboo Vision” in Malawi and the UK” [https://tvst.arvojournals.org/article.aspx?articleid=2720947] published. I am the First and Corresponding author.

I write to ask if I can re-use this paper as a chapter within my MD (Doctor of Medicine) thesis to the University of Glasgow. If so, thanks for letting me know any particular statements or comments that would be required (if any) with the thesis.

As some content will be covered in previous chapters of the MD thesis, the included content will require editing to avoid repetition. The formatting will also require amendment to suit the criteria stipulated by the University. Otherwise, the thrust of the content will be close to identical to the manuscript, including the Figures and Tables. All authors, their contribution, and the original published article, will all be fully cited.

Many thanks for letting me know

Appendix C: Normative Data from Teller Acuity Cards II. Extracts from Reference and Instruction Manual

	Patient Name: DOB: MR#:

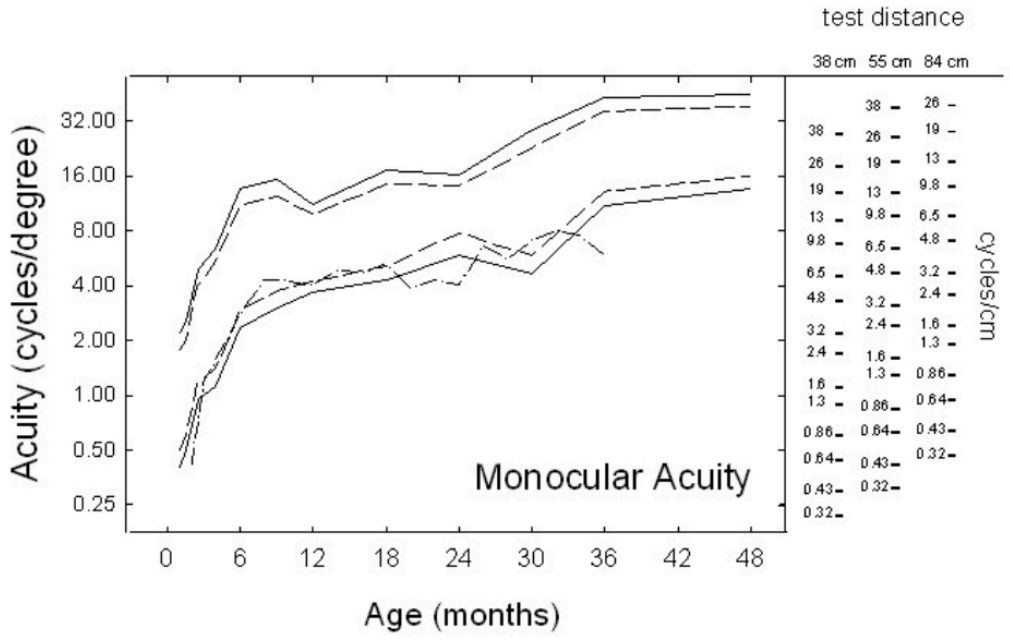
TELLER ACUITY CARDS® II - MONOCULAR AGE NORM CHART

Date_____/_____/_____ Gest. Age_____ (wks) Age_____(chron/corr)

Dx

RESULTS	card	cy/deg	Snellen	Diff between eyes
Both (Binoc)[sc/cc] [H/V]_____				☐ None ☐ 0.5 oct ☐ 1 oct ☐ >1 oct
Right eye (RE)[sc/cc] [H/V]_____				
Left eye (LE)[sc/cc] [H/V]_____				

COMMENTS Examiner_____



—·—·— Salomao & Ventura - 90% tolerance limits with 95% confidence, *Invest Ophthalmol Vis Sci*, 1995
——— Mayer et al. - 99% prediction limits, *Invest Ophthalmol Vis Sci*, 1995
– – – Mayer et al. - 95% prediction limits, *Invest Ophthalmol Vis Sci*, 1995

38

	Patient Name: Date of Birth: MR#:

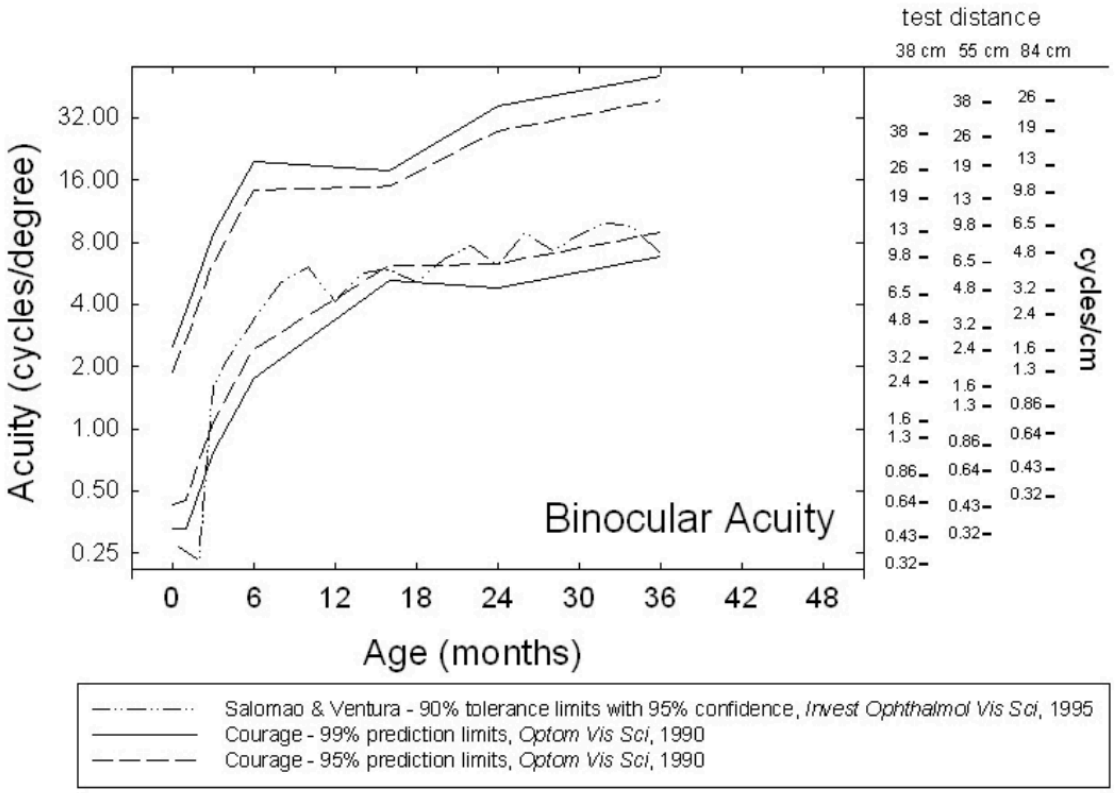
TELLER ACUITY CARDS® II - BINOCULAR AGE NORM CHART

Test Date_____/_____/_____ Gest. Age_____ (wks) Age_____ (chron/corr)

Dx

RESULTS	card	cy/deg	Snellen	Diff between eyes
Both (Binoc)[sc/cc] [H/V]_____				☐ None ☐ 0.5 oct ☐ 1 oct ☐ >1 oct
Right eye (RE) [sc/cc] [H/V]_____				
Left eye (LE)[sc/cc] [H/V]_____				

COMMENTS Examiner_____



37

Post Script

I feel extremely fortunate to have connected with inspirational individuals with unique skill sets that eclipse my own, who have helped progress this body of work, and continue to do so.

If you are considering an MD as a trainee, it is worth considering a few aspects. Complete it before you become a Consultant, and ideally before you have kids. The trade-off between completing the MD and the lost weekends required to get there, for me, became difficult to justify. At the time of writing, getting the Consultant Ophthalmologist position you want is unlikely to require an MD, so be clear about your reasons for doing it.

Consider a "staple job" MD by publication. 5 First Author publications is a reasonable target for the research time available for a 7-year run-through ophthalmology ST program. If you have the clarity and determination to ensure these papers form a cohesive narrative, it is possible to approach the University you have graduated from to set up a viva and obtain an MD. If you have led the work and gone through peer review for this number of papers, it is unlikely anyone on the panel will know the topic better than you. This option is less expensive, and if you are already publishing, significantly less work overall.

For me, the motivation for working in this area arises from disbelief regards the antiquated tools presently used for infant vision testing, and the sense that with the right team, we could improve what is available to identify poor vision in a young child at an earlier stage. While I think the paradigms developed are elegant and innovative, there has been negligible tangible change in over 30 years. I strongly believe that with available technology, infant vision testing can be entirely automated, more accurate, more comprehensive and delivered at a dramatically lower cost to a global community. The body of work captured in this thesis is a small step in that direction, I hope.

The soundtrack for this MD has predominantly been the music from Air, Zero 7, David Bowie, the Velvet Underground and Brian Eno.

Thanks for having a look at this book.

Iain