Preliminary Evaluation of a Smartphone App for Refractive Error Measurement

Gang Luo1,2, Chen-Yuan Lee1, Prerana Shivshanker1, Wenbo Cheng3, Jamie Wang4, Sophia Marusic5, Aparna Raghuram5,2, Yan Jiang6,2, and Rui Liu7

1 Schepens Eye Research Institute, Massachusetts Eye and Ear, Boston, MA, USA
2 Department of Ophthalmology, Harvard Medical School, Boston, MA, USA
3 The First Affiliated Hospital of Urumqi, Department of Ophthalmology, Xinjiang Medical University, Xinjiang, China
4 New England College of Optometry, Boston, MA, USA
5 Boston Children’s Hospital, Boston, MA, USA
6 Mass Eye and Ear Infirmary, Boston, MA, USA
7 Eye and ENT Hospital, NHC Key Laboratory of Myopia, Fudan University, Shanghai, China

Correspondence: Gang Luo, Schepens Eye Research Institute, Massachusetts Eye and Ear, 20 Stanford St., Boston, MA 02114, USA. e-mail: gangf.luo@meei.harvard.edu
Rui Liu, Eye & ENT Hospital, NHC Key Laboratory of Myopia, Fudan University, 83 Fenyang Road, Shanghai 200031, China. e-mail: lratb1@aliyun.com

Purpose: The purpose of this study was to evaluate the potential feasibility of using a smartphone app in myopia screening.

Methods: The app estimates myopic refractive error by measuring the far point distance for reading three 20/20 Tumbling E letters. In total, 113 myopic subjects with astigmatism no greater than $-1.75$ diopters (D) were enrolled from 5 sites. The mean age was 22 ± 8.5 years. The app measurement was compared with noncycloplegic subjective refraction measurement or autorefractor if subjective refraction was not available. In addition, 22 subjects were tested with the app for repeatability.

Results: For 201 eyes included, the range of spherical equivalent refraction error was 0 to $-10.2$ D. The app measurement and clinical measurement was highly correlated (Pearson $R = 0.91$, $P < 0.001$). There was a small bias (0.17 D) in the app measurement overall, and it was significantly different across the 5 sites due to different age of subjects enrolled at those sites ($P = 0.001$) – young adults in their 20s were underestimated the most by 0.49 D, whereas children were overestimated by 0.29 D. The mean absolute deviation of the app measurement was 0.65 D. The repeatability of multiple testing in terms of 95% limit of agreement was ±0.61 D.

Conclusions: Overall, the app measurement is consistent with clinical measurement performed by vision care professionals. The repeatability is comparable with that of some autorefractors. Age-associated human factors may influence the app measurement.

Translational Relevance: The app could be potentially used as a mass screening tool for myopia.

Introduction

Globally, myopia has reached epidemic proportions.1 In certain East Asian countries, more than half of the population is now myopic.2-4 By 2050, more than half of the world’s population is predicted to be myopic, and 10% of myopic individuals will have high myopia (refractive error ≤ −5 diopters [D]).5

This is a cause of concern because myopia is associated with an increased risk of glaucoma,6,7 cataract,8 and retinal detachment.9 Further, high myopia can lead to pathological myopia, which is associated with significant vision impairments not correctable with refraction and risk of permanent vision loss.10,11

Identifying early stage myopia in children and adolescents is a critical measure in mitigating the
myopia epidemic. This requires (1) the ability to screen a large number of children efficiently and (2) the ability to monitor myopia progression in a timely manner. The primary barriers to implementing myopia control measures are limited access to vision testing and limited eye care resources relative to the affected population of children. Mass screening of children is very challenging, especially for geographically large and diverse countries. Subjective professional refraction procedure for vision prescription is time-consuming and impractical for mass screening purposes. Utilizing autorefractors can more efficiently screen a given pool of potentially myopic individuals, but such programs can hardly cover all the population with the risk of myopia progression.

In addition, there is a huge need for measuring refractive error among adults in areas lacking vision care. Uncorrected refractive error (URE), although it is treatable, is the leading cause of visual impairment globally. About 116 million people are unnecessarily suffering from moderate and severe vision impairment due to URE. Studies have shown that the prevalence of URE is primarily driven by the low number of professionals and the lack of an adequate health infrastructure to make eye care obtainable in remote areas. For instance, in Sub-Saharan Africa, where there are only 2.5 eye doctors per million people (global mean is 31.7), the prevalence of blindness due to URE among adults aged above 50 years was estimated to be 10 times higher than that in high-income countries. Relying on the limited number of eye care professionals is not feasible to address the public health issue of URE in low-income countries and rural areas.

A strategy to address the mass screening problem may be recruiting nonprofessional personnel, such as school nurses, teachers, parents of students, and community healthcare workers. If they can be offered low cost, widely available, and validated tools for measuring refractive errors, they could make a strong working force to help accomplish the mass vision screening mission. Such an initiative was launched before using a low-cost refractometer for vision screening in refugees. Leveraging the latest mobile technologies in the digital era, we have developed a smartphone app to allow lay persons to measure refractive error, with minimal training and instruction. Compared to other mobile applications for refractive error measurement, such as Netra, a feature of our app is that it does not require any optical attachment, and uses the phone camera and screen only, making the app potentially more accessible simply through downloading. With a piece of software, there would be no issue related to hardware manufacturing, storage, distribution, and maintenance. Returning to the example of Sub-Saharan Africa, where disruptive vision care programs are needed, the region has about 237 million smartphone users in 2020. There are smartphone users in most communities. Ease of access to screening tools is crucial for successfully recruiting of many lay people to join the work force of mass screening in rural areas. Our standalone app can work without wireless data connection once it is installed via WiFi or wireless network. Considering the fact that low-cost Android phones are very popular in low- and middle-income countries, we have developed an Android version of the screening app, which was evaluated in this study. With these efforts, we try to create a tool suitable for hard-to-reach, low-income areas.

This study evaluates the app’s accuracy against conventional clinical methods of refraction measurement, as the first step to validate this smartphone app’s potential for mass screening.

Methods

Refraction App

The app estimates refractive error in terms of spherical equivalent (SE) diopter by measuring the far point for three 20/20 Tumbling E letters. In the course of measurement, the experimenter moves the phone screen toward the subject and verifies the responses. The angular size of the letter on the screen is maintained at 20/20 in real time based on the distance (Fig. 1), which is sensed using the selfie camera by the app based on facial features, such as interpupillary distance. If the subject can read at least two out of the
three randomized letters correctly, the testing for one eye is finished and the refractive error is calculated as the reciprocal of the distance. If the subject’s incorrect responses are more than one letter for a distance, the experimenter will move the phone screen closer to the subject and repeat the examination.

**Subjects**

Subjects were enrolled at 5 sites: (1) Massachusetts Eye and Ear Infirmary, (2) New England College of Optometry, (3) the First Hospital of Urumqi, (4) Shanghai Eye and ENT Hospital, and (5) Boston Children’s Hospital. Inclusion criteria were: 6 years and older, presence of myopia in at least one eye, astigmatism no greater than \(-1.75\) D, and no known ocular condition other than refractive error, such as cataract, macular degeneration, and glaucoma. In total, 201 eyes of 113 subjects were included. The mean age of our subjects was 22 ± 8.5 years ranging from 6 to 62 years and 67.2% were women. The races of the subjects included Hispanic (4%), Black (3%), Asian (60%), White (19%), and other/declined (13%). Out of the 113 subjects recruited for this study, 22 participated in a repeatability experiment, in which each eye of the enrolled 43 eyes was tested with the app 5 times. The mean age of these subjects was 22.7 ± 7.5 years, ranging from 7 to 35 years, and 63.6% were female subjects. In this study, the subjects were tested without eye dilation using the app and standard clinical methods, which included subjective refraction testing for 65.5% of subjects and autorefractor for the rest of the subjects.

This study received a single institutional review board (IRB) approval reviewed by Massachusetts Eye and Ear’s IRB for the three study sites in Boston, as well as the Shanghai ENT and the First Hospital of Urumqi’s IRBs. The study was conducted in accordance with the tenets of the Declaration of Helsinki.

**Testing Procedure**

After consent or parental consent were obtained, qualified clinicians performed standard noncycloplegic refraction. The measurement started with autorefractor or retinoscopy, followed by subjective refraction using phoropter/trial frame. Some participants (33 out of 113) did not receive subjective refraction because the testing could not be fit into the routine clinical procedure. For those participants, the autorefractor or retinoscopy measures were used. In this study, eyes with astigmatism more than \(-1.75\) D were excluded. This threshold was determined based on a pilot study, in which we observed that the 5% largest errors were associated with astigmatism more than \(-1.75\) D.

When using the app to measure refraction, the operator first placed the phone screen at least 2 meters away from the subject, to ensure the far point would not be missed. Starting from 2+ meters, the operator moved the phone toward the patients, at a speed gradually slowing down. If their moving speed was more than 0.5 D per second, a new, random set of letter stimuli would be shown, which reminds them to slow down. The testing was concluded if at least two out of three letters could be read correctly. Each eye was tested only once for comparison with standard clinical measurement.

**Statistical Analysis**

Linear regression and ANOVA were used to compare the app measurement against clinical measurement, and across different sites. SPSS software (version 24; IBM) was used for all statistical analyses.

**Results**

Based on standard clinical measurement, the SE refraction error in the 201 eyes ranged from 0 to \(-10.2\) D. The median (interquartile range [IQR]) was \(-2.88\) D (IQR = \(-1.63\) to \(-4.5\)), and 95% were below \(-6.75\) D. The IQR of astigmatism was 0 to \(-0.75\) D. In other words, 25% of eyes had astigmatism higher than \(-0.75\) D. Figure 2 shows the refraction error measure by the app as compared to the standard clinical measure. The two measures are highly correlated.
The differences between app and clinical measurements, aka, app measurement error, were different across sites. Positive values indicate underestimation. The mean age of participants for each site is shown on the plot. The sites are sorted in the order of mean age. Error bars represent standard error of mean. There seems to be a nonlinear pattern of age effect on the app measurement error.

($\text{Pearson } R = 0.91, P < 0.001$). Overall, the standard deviation of app measurement error was 0.83 D, and the mean absolute deviation (error) was 0.65 D.

The median IQR SE refraction error at sites 1 through 5 was (1) $-3.75$ D ($\text{IQR} = -2.75 \text{ to } -5.13$), (2) $-2.69$ D ($\text{IQR} = -1.03 \text{ to } -4.78$), (3) $-3.50$ D ($\text{IQR} = -2.53 \text{ to } -4.84$), (4) $-2.63$ D ($\text{IQR} = -1.50 \text{ to } -3.97$), and (5) $-2.34$ D ($\text{IQR} = -2.03 \text{ to } -3.44$), respectively. It was found by one-way ANOVA with Bonferroni correction that the app measurement error was significantly different across the sites ($P < 0.001, F_{4,197} = 6.1$). In Figure 3, the app measurement errors at the 5 sites are plotted in the order of mean age of the participants at those sites. The order was determined in the data analysis phase, not predetermined during subject recruitment. The difference in subject sources at those participating sites led to the difference in the age range. Figure 3 shows that the refraction was slightly overestimated for the oldest group at site 1 by 0.17 D and the youngest group at site 5 by 0.27 D, whereas the refraction was slightly underestimated for the rest of the participants at site 2 by 0.49 D, site 3 by 0.37 D, and site 4 by 0.17 D. As this seems to show a nonlinear pattern of age effect on the app measurement error, a univariate ANOVA was conducted with the participants from site 1 excluded ($n = 17$). Age and astigmatism were included as two covariates. It was found that age was a significant factor ($P = 0.001, F_{1,169} = 10.9$), but astigmatism was not ($P = 0.603, F_{1,169} = 0.27$).

The repeatability testing result based on 43 eyes is shown in Figure 4, in which each data point is the deviation of a measure from the average of repeated measures for each eye. The standard deviation of the variation was 0.31 D. The 95% of limits of agreement (i.e., 1.96 standard deviation) was ±0.61 D.

### Discussion and Conclusion

This preliminary study showed that myopic refractive error measurement with the mobile app was highly correlated with clinical standard measurement, up to $-10$ D (95% below $-6.75$ D), for eyes that did not have too strong astigmatism (less than 1.75 D). The app was used by different operators at 5 sites on subjects across a wide age range from 6 to 62 years. The overall bias was only 0.17 D underestimation, indicating the app measurement was quite accurate overall.

Although the bias varied across different sites, this difference was probably driven by the age of the participants at each site. Unlike autorefractor and retinoscope, which measure refraction objectively based on optical images, the app measures refraction by testing a subject's visual performance, which can be influenced by multiple factors rather than dictated by ocular optics alone. Different experience in vision testing, accommodation, and testing behavior related to age could impact visual performance in the app measurement. For instance, the participants at site 2 were all young (age 20s), experienced optometry students, who had participated in many practices and vision studies and therefore could guess the letter better than the other participants. Because the passing criterion was only two out of three letters, there may be a good chance the testing finished prematurely before the viewing distance reached the true far point. Consequently, their refraction error was underestimated the most. Similarly, the young adults at other sites might be underestimated to a certain extent because they had gone through vision tests since their young age.
Presumably, such an underestimation bias associated with experience can be eliminated by testing more trials. On the contrary, the participants at site 5 included mostly young minors, who can be accommodating variably to the target. Proximal accommodation (i.e., knowing that the object is close) and variable size of the target could influence ability to sustain clarity.\(^{21}\) Consequently, the far point they reached in the test might be relatively closer than that in standard conditions, in which the fogging technique in subjective refraction and landscape pictures in autorefractors are used to control accommodation. Thus, the app could overestimate their refraction error. Furthermore, the young minors, who are often less compliant than older patients, might delay in reporting when they saw the letters clearly. Hence the stopping distance would be shorter than the actual far point distance, resulting in overestimation. The speculations discussed above may explain the significant age effect found in this study when the data for young adults and minors are combined.

The oldest participant group at site 1 seemed to behave differently. Overall, they were overestimated by merely 0.17 D, which is smaller than prescription tolerance of 0.25 D. Unlike young participants in their 20s, they probably did not want to respond until they could see the letters with confidence. Thus, the testing was less likely to end prematurely, and the bias in their refraction measurement was very small. In addition, for participants older than 40 years of age, where accommodation loss starts to occur with age, the effect of changing size of the target would not have stimulated accommodation, as seen in the younger participants. Taken together, if the age effect found in this study is true, it probably could not be explained by a single mechanism.

Besides bias, another evaluation outcome from this study is the variability of the app measurement. Generally, the 0.84 D standard deviation of app measurement error mainly includes two components of variability—the repeatability of app measurement for the same eye and the intersubject variability, using sum of squares. The former was 0.31 D in this study, so it can be estimated that the app repeatability error contributed to 13.5% of the overall variability. Comparing to the 95% limit of agreement of some autorefractors, which were reported to be 0.21 D for cycloplegic autorefractor\(^{22}\) and about 0.70 D for noncycloplegic eyes,\(^{23,24}\) the app’s 95% limit of agreement (0.61 D) was comparable. If more trials are tested in a test, or multiple tests are conducted, the repeatability of average can be improved. For instance, four times of the test would improve the repeatability of mean by two times theoretically.

Excluding the intrasubject variability (app repeatability), the intersubject variability of the app measurement would be around 0.78 D. This could be related to various factors. This study investigated age and astigmatism; only age was found to be a significant intersubject factor. Although astigmatism should affect the performance of letter reading, and therefore could affect the refraction measurement, the enrollment criterion on astigmatism (better than 1.75 D) probably restricted its effect to a range so narrow that the small effect on the app measurement could not be found with our sample size. This suggests a favorable robustness of the app measurement. Because patients with high astigmatism are a small proportion in the children population for example, about 3.3% in 11 to 20-year-old children in Australia,\(^ {25}\) and 2.5% in 6 to 12-year-old children in Iran,\(^ {26}\) were greater than 2 D, 5.9% in 10 to 16-year-old children in China were greater than 1.5 D;\(^ {27}\) it is expected that the app can potentially measure spherical equivalent refraction error in the majority of the children population. Age is a risk factor of astigmatism. The prevalence of astigmatism greater than 2 D increases from about 3% in 20 to 30-year-old young adults to about 5% in 61 to 70-year-old people, and about 14% in the 71 to 80 year olds in Australia. As the increase in prevalence is not huge, it is expected that the app can also measure the majority of the adult population.

Another source of error may be due to the distance measurement method based on facial features. Because precise calibration for each individual was not included in the procedure for sake of simplicity, different facial features could result in an error in the far point distance estimation. In addition, as discussed above, the experience on reading letters near the acuity threshold could cause variability when the visual task is somewhat easy, especially for some experienced participants. If this was indeed one of the reasons, testing more trials may also help reduce the intersubject variability to a certain extent. It should be noted that, if the app is used for monitoring myopia progression longitudinally, the intersubject variability may not be a major concern.

In conclusion, considering the low bias and moderate variability, as well as the fact that the testing procedure is simple—moving the phone screen toward the patients until they just start to be able to discern the letters, it is possible for lay persons to administrate the refractive error measurement for the purpose of mass myopia screening. Large studies involving more novice operators and more diverse participants are warranted to further evaluate an improved version of the app.

There are still issues that need to be addressed before the smartphone can be actually utilized in mass screening. (1) The app can only measure myopic refractive
error. For patients with suspected hyperopia, a positive lens (e.g. +3 D) can be applied to them to artificially create myopia, and then measure with the app. The refraction offset can be subtracted from the measurement results. (2) The current version of the app is not able to measure the magnitude of astigmatism. For patients with strong astigmatism, the spherical equivalent results obtained by the app are very different from the clinical methods, which assess the spherical and cylinder power separately. If using a clock dial or a sunburst dial chart can confirm the presence of strong astigmatism, the patients can be flagged for further evaluation by professionals. (3) In mass screening, we also need to flag patients with ocular issues other than refractive error. As a commonly used and uncomplicated vision test, a pinhole visual acuity test can be administered as a prescreening examination. If one appears to have reduced vision (e.g. worse than 20/40), according to the pinhole test, the patient should be referred to specialists for a further examination. Although refraction does not need to be measured using the app for the patient, the pinhole testing result is still a meaningful outcome of the vision screening work. Certainly, the proposed solutions for these issues mentioned above would somewhat increase the procedure complexity of testing, because extra accessories and training are needed. How the increased complexity would impact the practical feasibility of mass screening using the mobile app needs to be further evaluated in future studies.

Acknowledgments

Supported in part with support from Harvard Catalyst | The Harvard Clinical and Translational Science Center (National Center for Advancing Translational Sciences, National Institutes of Health Award UL1 TR002541), and NIH T35EY007149-23. National Nature Science Foundation of China (82070997) and Shanghai Municipal Health Commission Program (201940485) supported this study as well. The funder had no role in study design, data collection and analysis, interpretation of data, or writing the manuscript.

Financial conflict interest: Gang Luo: pending patent. Gang Luo is a co-founder of EyeNexo, a startup company developing smartphone apps for vision tests. He did not participate in data collection. The company currently does not have any commercial product or license related to the subject matter of this article. No financial conflict interest exist for the other authors.

Disclosure: G. Luo, (P); C.-Y. Lee, None; P. Shivshanker, None; W. Cheng, None; J. Wang, None; S. Marusic, None; A. Raghuram, None; Y. Jiang, None; R. Liu, None

References

1. Morgan IG, French AN, Ashby RS, et al. The epidemics of myopia: Aetiology and prevention. Prog Retinal Eye Res. 2018;62:134–149.
2. Xu L, Zhuang Y, Zhang G, et al. Design, methodology, and baseline of whole city-million scale children and adolescents myopia survey (CAMS) in Wenzhou, China. Eye Vision (London, England). 2021;8:31.
3. Lee YY, Lo CT, Sheu SJ, et al. Risk factors for and progression of myopia in young Taiwanese men. Ophthalmic Epidemiol. 2015;22:66–73.
4. Quek TP, Chua CG, Chong CS, et al. Prevalence of refractive errors in teenage high school students in Singapore. Ophthalmic Physiol Opt. 2004;24:47–55.
5. Donovan LA, Sankaridurg P, Ho A, et al. Rates of Myopia Progression in Children. Invest Ophthalmol Vis Sci. 2010;51:1694–1694.
6. Mitchell P, Hourihan F, Sandbach J, et al. The relationship between glaucoma and myopia: the Blue Mountains Eye Study. Ophthalmology. 1999;106:2010–2015.
7. Marcus MW, de Vries MM, Junoy Montolio FG, et al. Myopia as a risk factor for open-angle glaucoma: a systematic review and meta-analysis. Ophthalmology. 2011;118:1989–1994.e1982.
8. Jeon S, Kim HS. Clinical characteristics and outcomes of cataract surgery in highly myopic Koreans. Korean J Ophthalmol: KJO. 2011;25:84–89.
9. Akiba J. Prevalence of Posterior Vitreous Detachment in High Myopia. Ophthalmology. 1993;100:1384–1388.
10. Fricke TR, Jong M, Naidoo KS, et al. Global prevalence of visual impairment associated with myopic macular degeneration and temporal trends from 2000 through 2050: systematic review, meta-analysis and modelling. Br J Ophthalmol. 2018;102:855–862.
11. Hayashi K, Ohno-Matsui K, Shimada N, et al. Long-term pattern of progression of myopic maculopathy: a natural history study. Ophthalmology. 2010;117:1595–1611, 1611.e1591-1594.
12. Saw S-M, Goh P-P, Cheng A, et al. Ethnicity-specific prevalences of refractive errors vary in
Asian children in neighbouring Malaysia and Singapore. Br J Ophthalmol. 2006;90:1230–1235.

13. Holden BA, Fricke TR, Ho SM, et al. Global Vision Impairment Due to Uncorrected Presbyopia. Arch Ophthalmol. 2008;126:1731–1739.

14. Flaxman SR, Bourne RRA, Resnikoff S, et al. Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis. Lancet Global Health. 2017;5:e1221–e1234.

15. Dean WH, Buchan JC, Gichuhi S, et al. Ophthalmology training in sub-Saharan Africa: a scoping review. Eye. 2021;35:1066–1083.

16. Steinmetz JD, Bourne RRA, Briant PS, et al. Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to Sight: an analysis for the Global Burden of Disease Study. Lancet Global Health. 2021;9:e144–e160.

17. Vincent JE. Simple Spectacles for Adult Refugees on the Thailand–Burma Border. Optom Vis Sci. 2006;83:803–810.

18. Tousignant B, Garceau M-C, Bouffard-Saint-Pierre N, et al. Comparing the Netra smartphone refractor to subjective refraction. Clin Exp Optometry. 2020;103:501–506.

19. Global System for Mobile Communications (GSMA). The Mobile Economy, Sub-Saharan Africa 2021. Available at: https://www.gsma.com/mobileeconomy/wp-content/uploads/2021/09/GSMA_ME_SSA_2021_English_Web_Singles.pdf.

20. Luo G. Measuring eye refraction. US Patent application 16/971,981, 2020.

21. Fогt N, Toole AJ, Rogers DL. A review of proximal inputs to the near response. Clin Exp Optom. 2016;99:30–38.

22. Moore KE, Berntsen DA. Central and Peripheral Autorefraction Repeatability in Normal Eyes. Optom Vis Sci. 2014;91:1106–1112.

23. Zadnik K, Mutti DO, Adams AJ. The repeatability of measurement of the ocular components. Invest Ophthalmol Vis Sci. 1992;33:2325–2333.

24. Bailey MD, Twa MD, Mitchell LG, et al. Repeatability of autorefraction and axial length measurements after laser in situ keratomileusis. J Cataract Refractive Surg. 2005;31:1025–1034.

25. Sanfilippo PG, Yazar S, Kearns L, et al. Distribution of astigmatism as a function of age in an Australian population. Acta Ophthalmologica. 2015;93:e377–e385.

26. Hashemi H, Asharlous A, Khabazkhoob M, et al. The profile of astigmatism in 6–12-year-old children in Iran. J Optom. 2021;14:58–68.

27. Li H, Li SM, Liu LR, et al. Astigmatism and its components in 12-year-old Chinese children: the Anyang Childhood Eye Study. Br J Ophthalmol. 2019;103:768–774.