



Evaluation of the test–retest and inter-mode comparability of the Impact of Vision Impairment questionnaire in people with chronic eye diseases

Jan Henrik Terheyden¹ · Reglind A. D. Ost¹ · Charlotte Behning² · Liza Mekschat¹ · Gamze Bildik¹ · Maximilian W. M. Wintergerst¹ · Frank G. Holz¹ · Robert P. Finger³

Received: 14 June 2023 / Revised: 3 November 2023 / Accepted: 30 November 2023
© The Author(s) 2024

Abstract

Purpose The main objective of this study is to assess the test–retest and inter-administration mode reliability of the Impact of Vision Impairment profile (IVI), a common patient-reported outcome measure (PROM) for people with chronic eye diseases.

Methods The IVI was administered to adult patients with stable, chronic eye diseases two to four times per participant (average intervals between administrations 12 to 20 days; maximum two phone interviews, paper administration, electronic administration) by two trained interviewers. Rasch models were fit to the data. Intra-class correlation coefficients (ICCs), mean differences and Cronbach's alpha between test–retest administrations (two phone interviews) and inter-mode comparisons were calculated.

Results Two hundred-sixteen patients (mean age 67 ± 12 years, 40% male) were included in the study. The IVI met all psychometric requirements of the Rasch model, and the division into the domains of functional items (IVI_F) and emotional items (IVI_E) corresponded to the German validation study. ICCs (all for IVI_F and IVI_E, respectively) for the retest administrations were 0.938 and 0.912, and 0.853 and 0.893 for inter-mode comparisons phone/paper, 0.939 and 0.930 for phone/electronic, and 0.937 and 0.920 for paper/electronic (all $p < 0.01$). Mean differences (all for IVI_F and IVI_E, respectively) for the retest administrations were 2.8% and 0.7% and ranged from 2.0% to 6.2% and from 0.4 % to 4.9% between administration modes. Cronbach's alpha ranged from 0.886 to 0.944 for retest and inter-mode comparisons.

Conclusion Due to the high test–retest reliability and the almost equally high comparability of different modes of administration of the IVI, the study endorses its use as a robust PROM to capture vision-related quality of life. Our results further support the use of the IVI as an endpoint in clinical trials and may simplify implementing it in both clinical trials or real-world evidence generation by offering multiple administration modes with high reliability.

Keywords Patient-reported outcome · Reliability · Impact of Vision Impairment scale · Quality of life

Jan Henrik Terheyden and Reglind A.D. Ost contributed equally to the work.

Extended author information available on the last page of the article

Key messages**What is known:**

- The Impact of Vision Impairment profile is a commonly used instrument for measuring vision-related quality of life
- Prior to this study, test-retest reliability data were only available from small cohorts

What is new:

- Our results suggest a high test-retest reliability in the current version of the IVI in a sample of >200 participants
- The data of our study demonstrate high comparability between three different modes of questionnaire administration

Introduction

Patient-reported outcome measures (PROMs) continue to gain importance in both routine eye care provision and research [1–4]. PROMs consist of several questions assessing the patient's health from their perspective [4]. One of the most common concepts measured by PROMs in ophthalmology is vision-related quality of life (VRQoL). It is known to be compromised by even mild forms of visual impairment and is therefore considered as a useful tool to gain additional information to visual function testing [5–7]. Typical PROMs in ophthalmology to assess VRQoL or visual functioning are the Impact of Vision Impairment (IVI) profile, the National Eye Institute Visual Functioning Questionnaire-25 (NEI-VFQ-25), and the Visual Function Index-14 (VF14) [4, 8, 9]. Similar to psychophysical function tests, PROM instruments require careful validation [10, 11]. One distinguishes between qualitative validation steps, where, e.g., content validity is assessed during the development phase of a questionnaire and quantitative validation steps [12]. The latter include an assessment of reliability and, e.g., construct validity of a questionnaire by statistical techniques such as exploratory factor analysis, to measure if the questions measure the construct they are intended to measure [12]. Similarly, reliability is assessed using quantitative metrics such as Cronbach's alpha or Rasch model person reliability to assess the internal consistency of the scale, and methods comparing two assessments with the same patient at different time points to examine the test–retest reliability of the scale [12].

Both the NEI-VFQ-25 and the VF14 have undergone test–retest reliability testing in large cohorts (NEI-VFQ-25 $n = 186$; VF-14 $n = 383$, which showed a high agreement between repeated administrations of both instruments [13, 14]. The IVI is widely used in rehabilitation and treatment studies [9, 11, 15–20], but to date, data on test–retest reliability have been generated in three studies that have included comparatively small subgroups of participants ($n \leq 60$ participants per study, total 102 test–retest participants in all available studies) [21–23]. Regarding this limited evidence,

we identified the need to reevaluate test–retest reliability in a larger cohort.

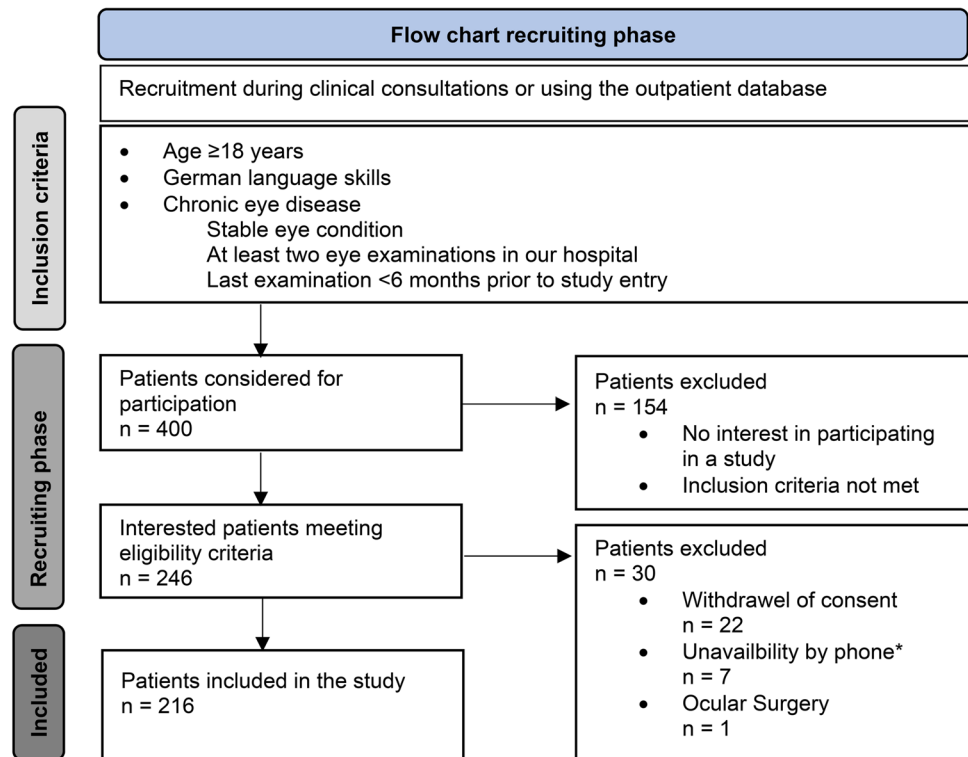
Another important aspect of questionnaire studies besides repeatability over different time points is the repeatability across different administration modes. Many questionnaire studies are still conducted with self-administered paper questionnaires or with interviews. However, due to numerous advantages of electronic modes of administration, such as higher and faster response rates, lower costs, and simplified data analysis, these have become more popular in research [24]. To our knowledge, a systematic comparison of electronic, paper-based, and interviewer-based modes has not been conducted in any of the studies mentioned above. One single study has compared self-administration of the IVI using a paper form and interviewer administration in a cohort with 31 participants, which is a similarly small sample size compared to the test–retest reliability studies described [23].

To fill this gap, our objective was to investigate the test–retest reliability and inter-mode comparability of the IVI concerning paper, phone, and electronic administration modes, hypothesizing that retest assessments are comparable but interviewer administration yielding higher IVI scores [25, 26].

Methods**Recruitment**

Our prospective study was carried out at the Department of Ophthalmology, University Hospital Bonn, Germany, from April 2020 to December 2020. Ethical approval was obtained from the Institutional Review Board of the University Hospital Bonn (reference number: 130/16). Patients were recruited during clinical consultations and from an outpatient database. Inclusion criteria were chronic eye conditions, age ≥ 18 years and sufficient German language skills. We only included individuals with stable chronic

Fig. 1 Flow chart of patient recruitment. *At least five phone call attempts



eye condition for at least two previous examinations at our hospital, the latter being no more than 6 months prior to study entry. Exclusion criteria were any acute-onset eye diseases < 6 weeks, any eye surgery or interventions during or < 6 weeks prior to study participation except intravitreal injections. The size of the sample was chosen according to international recommendations [27]. Written informed consent was obtained from all participants of the study. The study protocol followed the principles of the Declaration of Helsinki (Fig. 1).

The Impact of Vision Impairment questionnaire

The IVI questionnaire consists of 28 items to detect different aspects of VRQoL in patients with visual impairment [28]. It was originally developed to evaluate whether patients are limited in their daily lives due to reduced vision and may need rehabilitation [29, 30]. As indicated in its validation study, the German IVI includes four response options for items 1–13 and five response options for items 14–28, starting with “not at all” and ending with “very often” to evaluate the items’ influence on VRQoL. The fifth option “don’t do this for other reasons” was treated as missing in our analysis. The questionnaire is divided into two subscales: “Functional IVI” (items 1–20) and “Emotional IVI” (items 21–28) [5]. In accordance with the original IVI, the German version was initially developed for interview administration [5, 29].

Questionnaire administration

The study included two to four administrations of the IVI for each participant (Fig. 2). The initial mode of administration depended on the participant’s preference. The required response interval between administrations was ≤ 10 weeks. During the repeated administrations, the previous questionnaire responses were unavailable to both the interviewer and the participant. The interviewer administrations were conducted by two trained interviewers who instructed the participants according to recommended PROM administration guidelines [31].

In addition to the administration of the IVI, participants self-administered structured questionnaires on socio-demographical data and medical history. Diagnoses and visual acuity data from the last visit were obtained from the patient files. Regarding missing data, the questionnaire was excluded from the study if responses were available for fewer than 50% of items.

Psychometric validation of the IVI in our population

We performed Rasch analysis, which transforms the ordinal Likert scales into pseudo-interval-level scales (expressed in logits) and thus allows for parametric statistical analysis using commercial software (Winsteps software, ver. 3.92.1.2; Winsteps, Chicago, IL), employing the Andrich

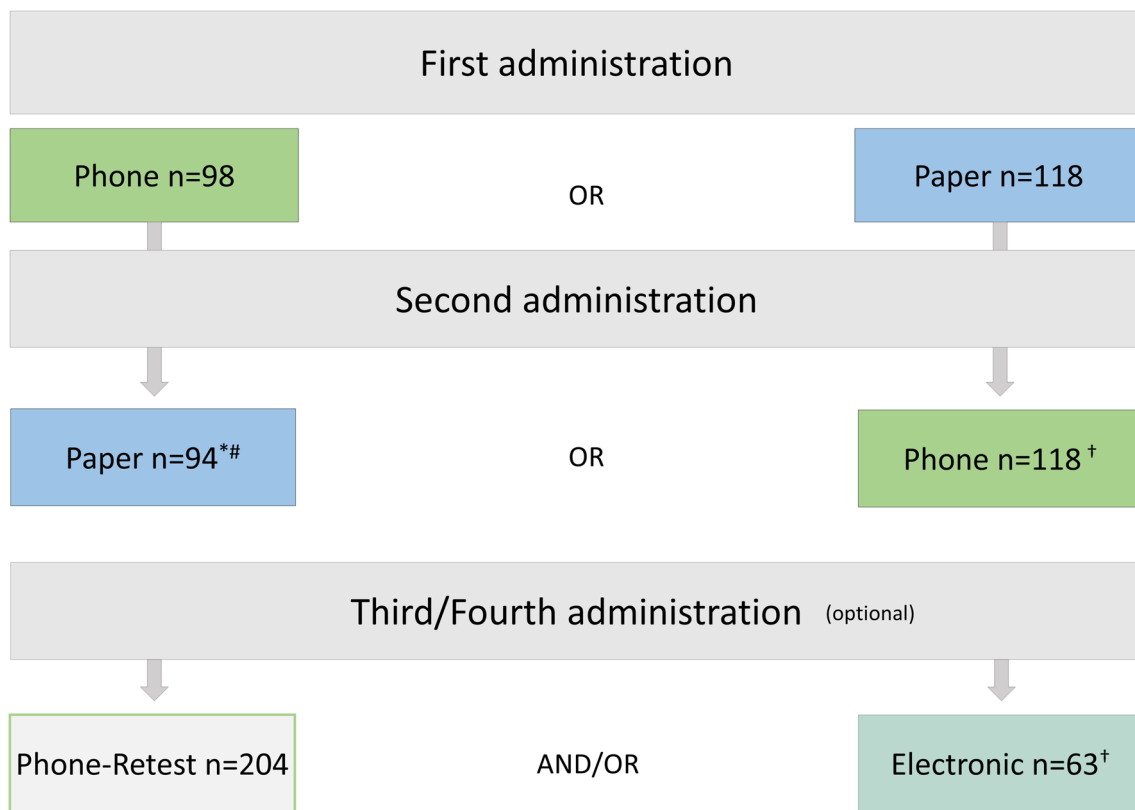


Fig. 2 Order of questionnaire administrations. Up to four sets of IVI questionnaire data were collected per participants. Interviewer administrations were performed remotely (phone) and self-administrations were performed using paper or electronic forms. *Omitted in 4 par-

ticipants due to vision impairment; #Performed as third mode by one participant; †Performed as third mode by two participants (after paper and electronic)

Rating Scale model [32, 33]. As an instance of latent trait models, the application of a Rasch model minimizes the impact of single missing responses since the measured construct is largely independent from individual missing values [34]. We anchored the person measures across modes of administration to the Rasch model based on the first interviewer administration to enable direct comparability of the results within our study, as well as in relation to other studies of the IVI [5, 35, 36]. Apart from this, we also analyzed item measures of Rasch models without this anchoring step, by Pearson correlation coefficients. After processing the questionnaire data, we inverted the person measures of the IVI to facilitate interpretation by assigning lower person measures to participants with more severe vision impairment.

Based on the Rasch models, we assessed threshold ordering, person reliability, person separation, item fit, and differential item functioning (DIF). Response category threshold ordering was assessed to verify whether the category structure belonging to the items is matching [28]. A person separation index (PSI) above 2.0 was considered as good and values between 1.5 and 2.0 as

acceptable, and a person reliability (PR) above 0.8 was deemed high and between 0.6 and 0.8 deemed acceptable [37–39]. Unidimensionality concerning the underlying construct was assessed using the infit and outfit mean square standardized residuals (MNSQ) statistic, with values up to 1.4 being reasonable for a rating scale survey [40]. The targeting of the instrument was identified by calculating the difference between person and item means, with values > 1.0 logits representing mistargeting. Lastly, differential item functioning (DIF) was assessed for each item to measure differences between subgroups concerning sex and age and was noticeable for values over 0.64 [41]. To determine the significance of deviating DIF values, a comparison of the person measures with and without retention of the conspicuous values was performed using intra-class correlation coefficients (ICCs).

Statistical analysis

The main outcomes of our study were ICCs and mean differences of repeated administration. We investigated

(1) test–retest and (2) inter-mode reliability using two-way mixed effect ICCs with absolute agreement as ICCs are the preferred assessment for settings where measures are repeated [42–44]. ICCs were interpreted according to the categories suggested by Cicchetti and Sparrow (poor $ICC < 0.40$; fair, $0.40 \leq ICC \leq 0.59$; good, $0.60 \leq ICC \leq 0.74$; excellent, $0.75 \leq ICC \leq 1.00$) [45]. Furthermore, we displayed the mean of inter-measurement differences ± 1.96 standard deviations as limits of agreement in Bland–Altman plots [46]. The coefficient of repeatability (CoR, i.e., an indicator of absolute reliability) was calculated as 1.96 standard deviations of the mean test–retest difference [47, 48]. Cronbach's alpha, a conventional test theory metric of reliability, was also calculated for the test–retest reliability assessment, based on person measures [12]. Deming regression analysis was performed to compare person measures between different administrations [49]. In contrast to classical linear regression, Deming regression considers the errors in both variables included in the respective regression model [50]. Intercept and slope were evaluated for significant deviations from 0 and 1, respectively [49]. To rule out any systematic bias, we compared the distribution of personal characteristics of participants who selected interviewer assessment and participants who selected self-assessment as the initial mode of administration, using the *t* test and the chi-squared test. In addition, we investigated which influence the initial mode of administration, visual acuity, self-reported hearing difficulties, self-reported psychiatric diseases, and the interval between administrations had on the mean differences between test–retest administration and phone-paper administration, using the Mann–Whitney *U* test.

Statistical analyses were performed with SPSS (Version 27 software, IBM SPSS Inc., Armonk, NY) and R (version 4.1.2). *P* values < 0.05 were considered statistically significant with correcting for multiple testing as necessary.

Results

Two hundred sixteen out of 246 total participants completed at least two administrations of the IVI and were included in the analysis. The reasons for drop-out were withdrawal of consent ($n = 22$), unavailability of participants by phone ($n = 7$), and ocular surgery ($n = 1$). The majority of participants (160 individuals, 74%) was above 60 years of age (Table 1). The most frequent ocular conditions in our cohort were vitreoretinal diseases, glaucoma and cataract (Supplementary Table 1). A total of 204 individuals were included in the test–retest assessment. The availability of administration modes for the inter-mode assessment was 216 participants (100%) for interviewer administrations, 212 (98%) for self-administrations using pen-and-paper,

Table 1 Characteristics of the sample

	<i>n</i> (%)
	All ($n = 216$)
Age	
Mean age [years] \pm SD	67 \pm 12
Sex	
Female (%)	129 (59.7)
Male (%)	87 (40.3)
Visual acuity (logMAR)	
Hearing difficulties	0.27 \pm 0.14
Yes (%)	59 (27.3)
No (%)	156 (72.2)
Missing (%)	1 (0.4)
Education	
Elementary school (%)	76 (35.2)
Secondary school (%)	78 (36.1)
High school (%)	22 (10.2)
University with graduation (%)	36 (16.7)
Missing (%)	4 (1.9)
Employment status	
Working (%)	56 (25.9)
Unemployed (%)	22 (10.2)
Retired (%)	130 (60.2)
Missing (%)	8 (3.7)
Living situation	
Alone (%)	64 (29.6)
With others (%)	148 (68.5)
Missing (%)	4 (1.9)
Marital status	
Married (%)	127 (58.8)
Widowed (%)	36 (16.7)
Divorced (%)	32 (14.8)
Unmarried (%)	21 (9.7)

and 63 (29%) for electronic self-administrations. The mean administration intervals were 18 ± 13 days for the test–retest assessments, 12 ± 12 days between phone and paper administrations, 20 ± 13 days between paper and electronic administrations, and 17 ± 11 days between phone and electronic administrations. The number of individuals who chose self-administration as the initial mode of administration ($n = 118$) was higher than the number of participants who chose to be interviewed first ($n = 98$) but no differences between these groups in terms of socio-demographic characteristics, ocular or systemic concurrent disease were present (all $p \geq 0.087$, Supplementary Table 1). Questionnaires with at least half of the responses missing were excluded from the study ($n = 5$). The number of single missing values amounted to 1.5%, or 0.01% of all answers, excluding the values of the answer option “Don't do this for other reasons”, which was considered missing in our analysis.

Table 2 Fit parameters of the phone, paper and electronic administration for the functional and emotional IVI

Parameters	Rasch model	Phone administration		Paper administration		Electronic administration	
		IVI_F	IVI_E	IVI_F	IVI_E	IVI_F	IVI_E
Threshold ordering, <i>n</i>	0	0	0	0	0	0	0
Misfitting items, <i>n</i>	0	0	item21^a	0	0	item1^a; item14^a	0
PSI	> 2.0 (1.5)	2.55	1.76	3.32	2.71	3.33	2.19
PR	> 0.8 (0.6)	0.87	0.76	0.92	0.88	0.92	0.83
Difference in person and item mean	< 1	1.91	2.60	1.56	2.87	2.38	3.97
DIF	< 0.64						
Age (≤ 60; > 60)		None	None	None	None	None	item28^b
Sex (female; male)		item1^b	None	item1^b; item9^b; item18^b	None	None	None

DIF, Differential item functioning; *IVI_F*, functional subscale of the IVI; *IVI_E*, emotional subscale of the IVI; *PR*, person reliability; *PSI*, person separation index

Bold values represent misfit to the Rasch model

^aValues are not degrading the measurement system [40]

^bNo influence on measurements ($ICC \geq 0.98$)

Rasch model fit

All response category thresholds were ordered. Three items demonstrated misfit (infit MNSQ 1.45, item 1; 1.56, item 14; 1.40, item 21), but the values were below a level that degrades the measurement system, and removal did not improve fit statistics, so we retained the items [40]. PR and PSI indicated adequate internal consistency (Table 2). The difference in person and item mean showed poor person-item targeting in our cohort. Three items were indicative of DIF by sex and one item by age group, but person measures were unchanged after removal, and the items could therefore be retained (IVI_F, phone administration, ICC 0.997, 95%-CI [0.996; 0.998]; IVI_F, paper administration, ICC 0.996, 95%-CI [0.995; 0.997]; IVI_E, electronic administration, ICC 0.979, 95%-CI [0.847; 0.993]). The resulting person measures were highest for the interview administered IVI (Supplementary Table 2). Pearson correlation coefficients between item measures were calculated to validate the chosen anchoring method. They were 0.975 [0.946; 0.988] for the test and retest-phone administrations, 0.878 [0.751; 0.942] for phone and paper administration, and 0.877 [0.748; 0.942] for phone and electronic administration.

Test–retest and inter-mode reliability

ICCs were excellent, both for test–retest and inter-mode comparisons (Table 3). Mean differences between test–retest and inter-mode comparisons were comparable, and the Bland–Altman analysis did not indicate any systematic bias (Fig. 3). Deming regression revealed that absolute differences of test–retest assessments were not significantly

different from each other when the same mode of administration was used. Phone administration compared to paper and electronic administration resulted in significantly higher functional IVI subscale scores (i.e., higher VRQoL), and paper administration compared to phone and electronic administration yielded significantly lower emotional IVI subscale scores (i.e., lower VRQoL) when comparing the Deming regression intercepts (Table 3; Supplementary Table 2).

Sensitivity analysis

To validate the above findings, we investigated relationships between mean differences of the test–retest or inter-mode comparisons and five potential confounders, correcting for multiple testing with the Bonferroni method. Neither best-corrected visual acuity (corrected $p \geq 0.546$), nor psychiatric diseases (corrected $p \geq 1.0$), self-reported hearing difficulties (corrected $p \geq 1.0$), administration interval (corrected $p \geq 1.0$), nor the initial mode of administration (corrected $p \geq 1.0$) were significantly associated with the IVI subscale score differences (Supplementary Table 3).

Discussion

The purpose of this study was to evaluate test–retest reliability and inter-mode reliability. Test–retest differences were small and non-significant, and inter-mode variations were in a similar, small range. We found a small but significant reduction in VRQoL scores when questionnaires were self-administered using paper forms compared to phone administration. Due

Table 3 Reliability metrics of the test–retest and inter-mode assessments for the functional and emotional IVI subscales

		Test–retest		Inter-mode	
		<i>Phone–phone</i> (n = 204)	<i>Phone–paper</i> (n = 212)	<i>Phone–electronic</i> (n = 63)	<i>Paper–electronic</i> (n = 63)
IVI_F	ICC (average values)	0.938	0.853	0.939	0.937
	95%-CI	[0.909; 0.957]	[0.693; 0.917]	[0.896; 0.964]	[0.894; 0.962]
	Cronbach's α	0.94	0.89	0.94	0.94
	Mean difference (% subscale range)	0.27 (2.8)	0.61 (6.2)	0.24 (2.5)	–0.20 (2.0)
	CoR (% subscale range)	1.55 (15.8)	2.06 (21.0)	1.61 (16.5)	1.53 (15.6)
	Deming, intercept	–0.10	–0.29	–0.27	–0.18
	95%-CI	[–0.27; 0.07]	[–0.53; –0.06]*	[–0.54; –0.01]*	[–0.43; 0.07]
	Deming, slope	0.91	0.83	1.02	1.24
	95%-CI	[0.82; 0.99]*	[0.72; 0.94]*	[0.86; 1.17]	[1.06; 1.42]*
	IVI_E	ICC (average values)	0.912	0.893	0.930
95%-CI		[0.884; 0.933]	[0.816; 0.933]	[0.885; 0.958]	[0.866; 0.952]
Cronbach's α		0.91	0.91	0.93	0.92
Mean difference (% subscale range)		–0.06 (0.7)	0.41 (4.9)	0.03 (0.4)	–0.20 (2.3)
CoR (% subscale range)		1.77 (21.2)	1.75 (20.9)	1.42 (16.9)	1.40 (16.8)
Deming, intercept		0.01	–0.38	0.25	0.35
95%-CI		[–0.27; 0.28]	[–0.60; –0.16]*	[–0.17; 0.67]	[0.16; 0.54]*
Deming, slope		1.03	0.99	0.83	0.89
95%-CI		[0.90; 1.16]	[0.88; 1.09]	[0.62; 1.04]	[0.76; 1.03]

ICC, Intra-class correlation coefficient; IVI_F, functional subscale; IVI_E, emotional subscale; ICC, intra-class correlation coefficient; CoR, coefficient of repeatability

*Regression intercept significantly different from 0 or slope significantly different from 1

to the excellent ICC values (IVI_F 0.853; IVI_E 0.893) and low mean differences (IVI_F 6% of subscale range, IVI_E 5% of subscale range), we classified these deviations in the statistically but not clinically relevant range [45]. Overall, our data demonstrate that the IVI questionnaire is highly reliable, independent of repeated assessment or the mode of administration, also supported by the strong associations of the item measures across different modes. Our results were not affected by the initial mode of administration, participants' visual or hearing impairment, psychiatric diseases, or the time between IVI administrations.

The test–retest reliability of the IVI was largely comparable to what has been reported previously in smaller cohorts (n = 20 to 60) [21–23]. ICCs between IVI test–retest scores in our sample (0.94 for IVI_F and 0.91 for IVI_E) were minimally higher compared to the values reported by the original 32-item IVI by Weih et al. (0.88 for IVI total sum score) and mean differences were lower in our study ($\leq 2.8\%$ of subscale range versus 6.1% of scale range) [23]. Test–retest subgroup analyses in the Greek and Thai validations of the IVI reported similar results to our study (ICCs ≥ 0.90 , Cronbach's $\alpha > 0.75$) [21, 22]. Other commonly used ophthalmic PROMs, such as the NEI-VFQ-25 or the VF-14, achieve

similarly ICCs (0.57–0.91), which places the IVI in the series of reliable PROMs in ophthalmology [13, 14, 51].

The implementation of different modes of administration did not have a relevantly higher impact on the IVI subscale scores than performing a retest administration in our study. As indicated above, the inter-mode reliability has only been investigated in one study (n = 31) that compared self-administration to interview-administration, using IVI-32 sum scores which are no longer recommended to be used [23]. In contrast to Weih et al. we have investigated the inter-mode reliability of the IVI using more state-of-the-art statistical approaches including Rasch-based person measures. Our results suggested a higher agreement between modes of administration than previously reported (mean differences $\leq 6.2\%$ in this study versus $\leq 9.5\%$ in the previous analysis) [23]. Additionally, our findings of the electronic questionnaire with excellent ICCs ≥ 0.920 further support the implementation of electronic PROMs in ophthalmology (Table 3) [24].

We found a trend of higher reported vision-related quality of life in interviews compared to pen and paper assessments, which is known from other PROMs such as the vision core measure (VCM1) study by Frost et al. [25]. In our dataset,

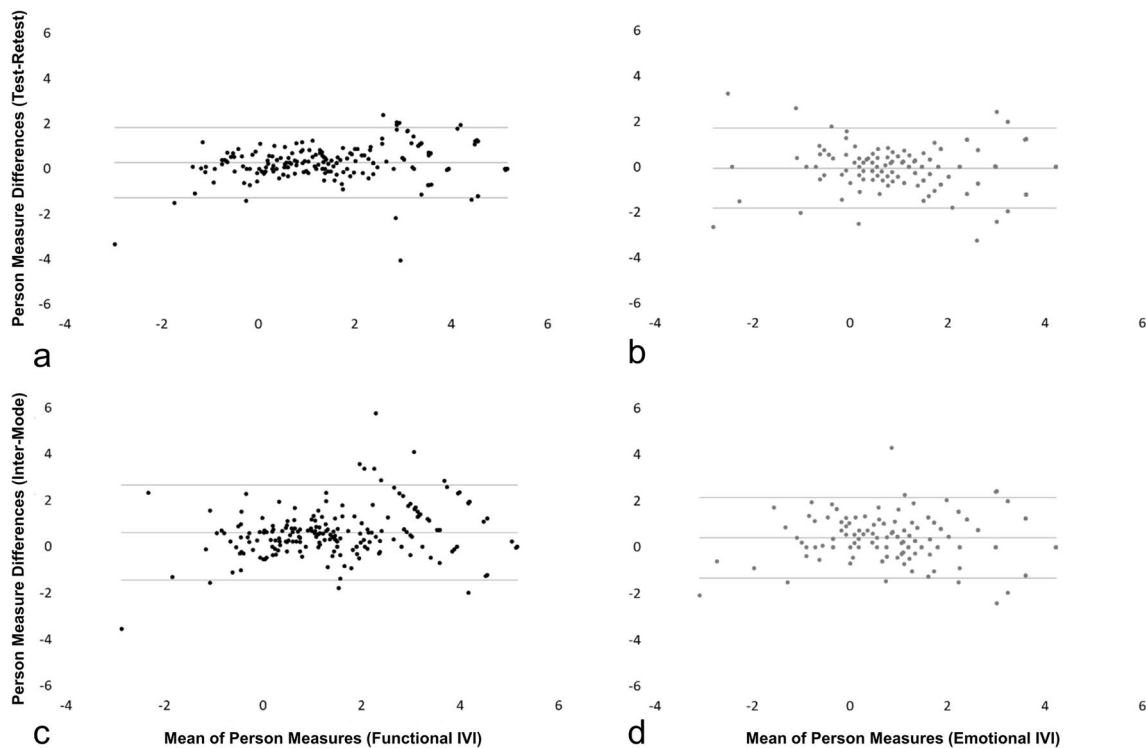


Fig. 3 Bland–Altman plots of test–retest assessments (phone administration) of the functional (a) and emotional (b) subscales and of inter-mode comparability assessments (phone and paper administra-

tion) of the functional (c) and emotional (d) IVI subscales display the distributions of retest differences in person measures (higher vision-related quality of life corresponds to higher person measures)

mean differences between different modes of administration were small ($\leq 6.2\%$ of functional subscale range; $\leq 4.9\%$ of emotional subscale range), further supporting the equivalence between the investigated administration modes for the IVI. When comparing paper to electronic self-administration, mean differences were $\leq 2.3\%$ of the respective subscale range. This is in line with previous research, which has identified 5% of the scale range as the cutoff value for most studies comparing paper to electronic administrations of various PROMs [52]. Only few other studies have systematically compared the use of different modes of administration in ophthalmic PROMs, and reported inter-mode reliabilities were in an overall similar range to the results of our study [53, 54].

The main limitation of our study is the targeting of the sample which is a result of including individuals with several chronic eye diseases but not only visually impaired individuals [55]. The problem of inadequate person-item targeting is known from previous studies of the IVI [56]. We did not randomize participants by initial mode of administration since most participants asked to start with pen and paper administration. Yet, a post hoc analysis did not reveal any differences based on the initial mode of administration. An additional aspect not investigated in our study is how the participants' health literacy may

modify the reliability of the IVI, which should be further investigated in future studies [57].

Our study's main strengths include its large sample size in which both test–retest and inter-mode reliability of the IVI profile have been tested, the heterogeneity of the sample making the results more likely to be generalizable to a population accessing eye care services, and the use of latent trait models which have several known advantages over sum scoring in psychometric assessment [58, 59].

In conclusion, the IVI questionnaire demonstrated excellent test–retest reliability and our data suggest the use of interview, paper-based, or electronic modes of administration to be comparable in our study cohort. The results may facilitate implementing the IVI in clinical routine and research because of a reduced administration burden.

Abbreviations *CoR*: Coefficient of repeatability; *DIF*: Differential item functioning; *ICC*: Intra-class correlation coefficient; *IVI*: Impact of Vision Impairment profile; *IVI_F*: Functional subscale of the Impact of Vision Impairment profile; *IVI_E*: Emotional subscale of the Impact of Vision Impairment profile; *MNSQ*: Mean square standardized residuals; *PR*: Person reliability; *PROMs*: Patient-reported outcome measures; *PSI*: Person separation index; *VRQoL*: Vision-related quality of life

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00417-023-06334-4>.

Author contribution JHT, RADO, and RPF designed the study. RADO, LM, and GB collected the data. JHT, RADO, CB, RPF, MWMW, and FGH analyzed and interpreted the data. JHT and RADO were main contributors in writing the manuscript. All authors read, substantively revised, and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethical approval Ethical approval was obtained from the Institutional Review Board of the University Hospital Bonn (reference number: 130/16). The study protocol followed the principles of the Declaration of Helsinki.

Informed consent Written informed consent was obtained from all participants of the study.

Consent for publication Not applicable.

Competing interests JHT: Heidelberg Engineering, Optos, Zeiss, CenterVue, Novartis, Okko; RADO: Heidelberg Engineering, Optos, Zeiss, CenterVue; CB: none; LM: Heidelberg Engineering, Optos, Zeiss, CenterVue; GB: Heidelberg Engineering, Optos, Zeiss, CenterVue; MWMW: Heidelberg Engineering (F, R), Optos (F), Carl Zeiss Meditec (F), CenterVue (F), Heine Optotechnik GmbH (C, R, F), Berlin-Chemie AG (F), Novartis Pharma GmbH (R, F), D-Eye Srl (F), Eyenuk Inc. (F), ASKIN & CO GmbH (R), DigiSight Technologies (R); FGH: Acucela, Allergan, Apellis, Bayer, Boehringer-Ingelheim, Bioeq/Formycon, CenterVue, Ellex, Roche/Genentech, Geuder, Grayburg Vision, Heidelberg Engineering, Kanghong, LinBioscience, NightStarX, Novartis, Optos, Pixium Vision, Oxurion, Stealth Bio-Therapeutics, Zeiss; RPF: Bayer, Ellex, Novartis, Novartis, Opthea, Alimera, Santhera, Roche/Genentech, CentreVue, Zeiss.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.




References

- Loukanova S, Bridges J (2008) Empowerment in medicine: an analysis of publication trends 1980–2005. *Open Med* 3:105–110. <https://doi.org/10.2478/s11536-007-0066-7>
- Nelson EC, Eftimovska E, Lind C et al (2015) Patient reported outcome measures in practice. *BMJ* 350:g7818. <https://doi.org/10.1136/bmj.g7818>
- Dean S, Mathers JM, Calvert M et al (2017) “The patient is speaking”: discovering the patient voice in ophthalmology. *Br J Ophthalmol* 101:700–708. <https://doi.org/10.1136/bjophthalmol-2016-309955>
- Braithwaite T, Calvert M, Gray A et al (2019) The use of patient-reported outcome research in modern ophthalmology: impact on clinical trials and routine clinical practice. *Patient Relat Outcome Meas* 10:9–24. <https://doi.org/10.2147/PROM.S162802>
- Finger RP, Fenwick E, Marella M et al (2011) The Impact of Vision Impairment on vision-specific quality of life in Germany. *Invest Ophthalmol Vis Sci* 52:3613–3619. <https://doi.org/10.1167/iovs.10-7127>
- Varma R, Wu J, Chong K et al (2006) Impact of severity and bilaterality of visual impairment on health-related quality of life. *Ophthalmology* 113:1846–1853. <https://doi.org/10.1016/j.ophtha.2006.04.028>
- Knauer C, Pfeiffer N (2008) The value of vision. *Graefes Arch Clin Exp Ophthalmol* 246:477–482. <https://doi.org/10.1007/s00417-007-0668-4>
- de Boer MR, Pluijm SMF, Lips P et al (2004) Different aspects of visual impairment as risk factors for falls and fractures in older men and women. *J Bone Miner Res* 19:1539–1547. <https://doi.org/10.1359/JBMR.040504>
- Luu W, Kalloniatis M, Bartley E et al (2020) A holistic model of low vision care for improving vision-related quality of life. *Clin Exp Optom* 103:733–741. <https://doi.org/10.1111/cxo.13054>
- Massof RW, Rubin GS (2001) Visual function assessment questionnaires. *Surv Ophthalmol* 45:531–548. [https://doi.org/10.1016/S0039-6257\(01\)00194-1](https://doi.org/10.1016/S0039-6257(01)00194-1)
- Lamoureux EL, Pallant JF, Pesudovs K et al (2007) The effectiveness of low-vision rehabilitation on participation in daily living and quality of life. *Invest Ophthalmol Vis Sci* 48:1476–1482. <https://doi.org/10.1167/IOVS.06-0610>
- Frost MH, Reeve BB, Liepa AM et al (2007) What is sufficient evidence for the reliability and validity of patient-reported outcome measures? *Value Health* 10(Suppl 2):94–105. <https://doi.org/10.1111/j.1524-4733.2007.00272.x>
- Labiris G, Katsanos A, Fanariotis M et al (2008) Psychometric properties of the Greek version of the NEI-VFQ 25. *BMC Ophthalmol* 8:4. <https://doi.org/10.1186/1471-2415-8-4>
- Cassard SD, Patrick DL, Damiano AM et al (1995) Reproducibility and responsiveness of the VF-14. An index of functional impairment in patients with cataracts. *Arch Ophthalmol* 113:1508–1513. <https://doi.org/10.1001/archophth.1995.01100120038005>
- Hassell JB, Lamoureux EL, Keeffe JE (2006) Impact of age related macular degeneration on quality of life. *Br J Ophthalmol* 90:593–596. <https://doi.org/10.1136/bjo.2005.086595>
- Pondorfer SG, Terheyden JH, Heinemann M et al (2019) Association of vision-related quality of life with visual function in age-related macular degeneration. *Sci Rep* 9:15326. <https://doi.org/10.1038/s41598-019-51769-7>
- Tan JCK, Nguyen V, Fenwick E et al (2019) Vision-related quality of life in keratoconus: a save sight keratoconus registry study. *Cornea* 38:600–604. <https://doi.org/10.1097/ICO.0000000000001899>
- Finger RP, Guymer RH, Gillies MC et al (2014) The impact of anti-vascular endothelial growth factor treatment on quality of life in neovascular age-related macular degeneration. *Ophthalmology* 121:1246–1251. <https://doi.org/10.1016/j.ophtha.2013.12.032>
- Kanellopoulos AJ (2019) The impact of keratoconus treatment with the Athens Protocol (partial topography-guided photorefractive keratectomy combined with higher-fluence corneal collagen cross-linking) on quality of life: a long-term study. *Clin Ophthalmol* 13:795–803. <https://doi.org/10.2147/OPHT.S188519>
- de Boer MR, Moll AC, de Vet HCW et al (2004) Psychometric properties of vision-related quality of life questionnaires: a systematic review. *Ophthalmic Physiol Opt* 24:257–273. <https://doi.org/10.1111/j.1475-1313.2004.00187.x>
- Marakis TP, Koutsandrea C, Poulou MS (2020) The Impact of Vision Impairment on vision-related quality of life of patients with neovascular age-related macular degeneration. *Eur J Ophthalmol* 32:481–490. <https://doi.org/10.1177/1120672120972625>

22. Ratanasukon M, Tongsomboon J, Bhurayanontachai P et al (2016) The Impact of Vision Impairment (IVI) questionnaire; validation of the Thai-version and the implementation on vision-related quality of life in Thai rural community. *PLoS ONE* 11:e0155509. <https://doi.org/10.1371/journal.pone.0155509>
23. Weih LM, Hassell JB, Keeffe J (2002) Assessment of the Impact of Vision Impairment. *Invest Ophthalmol Vis Sci* 43:927–935
24. Evans JR, Mathur A (2005) The value of online surveys. *Internet Res* 15:195–219. <https://doi.org/10.1108/10662240510590360>
25. Frost NA, Sparrow JM, Hopper CD et al (2001) Reliability of the VCM1 questionnaire when administered by post and by telephone. *Ophthalmic Epidemiol* 8:1–11. <https://doi.org/10.1076/opep.8.1.1.1539>
26. Wolffsohn JS, Cochrane AL, Watt NA (2000) Implementation methods for vision related quality of life questionnaires. *Br J Ophthalmol* 84:1035–1040. <https://doi.org/10.1136/bjo.84.9.1035>
27. Mokkink LB, Prinsen CAC, Patrick DL et al. (2019) COSMIN study design checklist for patient-reported outcome measurement instruments. https://www.cosmin.nl/wp-content/uploads/COSMIN-study-designing-checklist_final.pdf. Accessed 10 Oct 2023
28. Lamoureux EL, Pallant JF, Pesudovs K et al (2006) The Impact of Vision Impairment questionnaire: an evaluation of its measurement properties using Rasch analysis. *Invest Ophthalmol Vis Sci* 47:4732–4741. <https://doi.org/10.1167/IOVS.06-0220>
29. Keeffe JE, McCarty CA, Hassell JB et al (1999) Description and measurement of handicap caused by vision impairment. *Aust N Z J Ophthalmol* 27:184–186. <https://doi.org/10.1046/j.1440-1606.1999.00179.x>
30. Hassell JB, Weih LM, Keeffe JE (2000) A measure of handicap for low vision rehabilitation: the Impact of Vision Impairment profile. *Clin Exp Ophthalmol* 28:156–161. <https://doi.org/10.1046/j.1442-9071.2000.00312.x>
31. Terheyden JH, Ponderer SG, Behning C et al (2023) Disease-specific assessment of vision impairment in low luminance in age-related macular degeneration - a MACUSTAR study report. *Br J Ophthalmol* 107:1144–1150. <https://doi.org/10.1136/bjophthalmol-2021-320848>
32. Wright BD, Linacre JM (1989) Observations are always ordinal; measurements, however, must be interval. *Arch Phys Med Rehabil* 70:857–860
33. Linacre JM (2006) A user's guide to Winsteps/Ministep: Rasch-model computer programs. Mesa Press, Chicago
34. Waterbury GT (2019) Missing data and the Rasch model: the effects of missing data mechanisms on item parameter estimation. *J Appl Meas* 20:154–166
35. Herzog AR, Rodgers WL (1988) Interviewing older adults: mode comparison using data from a face-to-face survey and a telephone resurvey. *Public Opin Q* 52:84–99
36. Hochstim JR (1967) A critical comparison of three strategies of collecting data from households. *J Am Stat Assoc* 62:976. <https://doi.org/10.2307/2283686>
37. Wright BD, Masters GN (1982) Rating scale analysis: Rasch measurement. Mesa Press, Chicago
38. Boone WJ, Noltemeyer A (2017) Rasch analysis: a primer for school psychology researchers and practitioners. *Cogent Educ* 4:202. <https://doi.org/10.1080/2331186X.2017.1416898>
39. Bond TG, Fox CM (2001) Applying the Rasch model: fundamental measurement in the human sciences, 1st edn. Lawrence Erlbaum Associates Publishers, New York
40. Wright BD, Linacre JM (1994) Reasonable mean-square fit values. *Rasch Measure Transact* 8:370–371
41. Zwick R, Thayer DT, Lewis C (1999) An empirical bayes approach to Mantel-Haenszel DIF Analysis. *J Educ Meas* 36:1–28. <https://doi.org/10.1111/j.1745-3984.1999.tb00543.x>
42. Mohamad Adam Bujang, Hon Yoon Khee, Lee Keng Yee (2022) A step-by-step guide to questionnaire validation research. <https://doi.org/10.5281/zenodo.6801208>
43. McGraw KO, Wong SP (1996) Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1:30–46. <https://doi.org/10.1037/1082-989X.1.1.30>
44. Koo TK, Li MY (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 15:155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
45. Cicchetti DV, Sparrow SA (1981) Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *Am J Ment Defic* 86:127–137
46. Giavarina D (2015) Understanding Bland Altman analysis. *Biochem Med (Zagreb)* 25:141–151. <https://doi.org/10.11613/BM.2015.015>
47. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1:307–310
48. Vaz S, Falkmer T, Passmore AE et al (2013) The case for using the repeatability coefficient when calculating test-retest reliability. *PLoS ONE* 8:e73990. <https://doi.org/10.1371/journal.pone.0073990>
49. Deming WE (1943) Statistical adjustment of data. J Wiley & Sons Inc, New York
50. Ciccione L, Dehaene S (2021) Can humans perform mental regression on a graph? Accuracy and bias in the perception of scatterplots. *Cogn Psychol* 128:101406. <https://doi.org/10.1016/j.cogpsych.2021.101406>
51. Nichols KK, Mitchell GL, Zadnik K (2002) Performance and repeatability of the NEI-VFQ-25 in patients with dry eye. *Cornea* 21:578–583. <https://doi.org/10.1097/00003226-200208000-00009>
52. Gwaltney CJ, Shields AL, Shiffman S (2008) Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review. *Value Health* 11:322–333. <https://doi.org/10.1111/j.1524-4733.2007.00231.x>
53. Clayton JA, Eydelman M, Vitale S et al (2013) Web-based versus paper administration of common ophthalmic questionnaires: comparison of subscale scores. *Ophthalmology* 120:2151–2159. <https://doi.org/10.1016/j.ophtha.2013.03.019>
54. Terheyden JH, Mekschat R, Ost RAD et al (2022) Interviewer administration corresponds to self-administration of the vision impairment in low luminance (VILL) questionnaire. *Trans Vis Sci Tech* 11:21. <https://doi.org/10.1167/tvst.11.4.21>
55. Goldstein JE, Fenwick E, Finger RP et al (2018) Calibrating the Impact of Vision Impairment (IVI): creation of a sample-independent visual function measure for patient-centered outcomes research. *Translat Vision Sci Technol* 7:38. <https://doi.org/10.1167/tvst.7.6.38>
56. Lamoureux EL, Ferraro JG, Pallant JF et al (2007) Are standard instruments valid for the assessment of quality of life and symptoms in glaucoma? *Optom Vis Sci* 84:789–796. <https://doi.org/10.1097/OPX.0b013e3181334b83>
57. Taylor DJ, Jones L, Edwards L et al (2021) Patient-reported outcome measures in ophthalmology: too difficult to read? *BMJ Open Ophthalmol* 6:e000693. <https://doi.org/10.1136/bmjophth-2020-000693>
58. Ayton LN, Rizzo JF, Bailey IL et al (2020) Harmonization of outcomes and vision endpoints in vision restoration trials: recommendations from the international HOVER taskforce. *Trans Vis Sci Tech* 9:25. <https://doi.org/10.1167/tvst.9.8.25>
59. Hobart J, Cano S (2009) Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technol Assess* 13:iii, ix–x 1–177. <https://doi.org/10.3310/hta13120>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Jan Henrik Terheyden¹  · Reglind A. D. Ost¹ · Charlotte Behning² · Liza Mekschat¹ · Gamze Bildik¹ · Maximilian W. M. Wintergerst¹  · Frank G. Holz¹ · Robert P. Finger³ 

✉ Jan Henrik Terheyden
jan.terheyden@ukbonn.de

✉ Robert P. Finger
RobertPatrick.Finger@umm.de

¹ Department of Ophthalmology, University Hospital Bonn,
Ernst-Abbe-Str. 2, 53127 Bonn, Germany

² Department of Medical Biometry, Informatics
and Epidemiology, University Hospital Bonn, Bonn,
Germany

³ Department of Ophthalmology, University Hospital
Mannheim, University of Heidelberg, Mannheim, Germany