



Clinical Neuroscience

Reliability of visual orienting response measures in children with and without visual impairments



Marlou J.G. Kooiker^{a,*}, Johannes van der Steen^{a,b}, Johan J.M. Pel^a

^a Vestibular and Ocular Motor Research Group, Department of Neuroscience, Erasmus MC, PO Box 2040, 3000 CA Rotterdam, The Netherlands

^b Royal Dutch Visio, Huizen, The Netherlands

HIGHLIGHTS

- Orienting responses are a quantitative measure of visual information processing.
- Response times and fixation accuracies are stable over time in young children.
- The method has good test–retest reliability in children with visual impairments.
- The orienting eye movement responses provide a criterion for reliable data.

ARTICLE INFO

Article history:

Received 27 February 2014

Received in revised form 4 June 2014

Accepted 5 June 2014

Available online 12 June 2014

Keywords:

Visual orienting responses

Remote eye tracking

Visual information processing

Test–retest reliability

Infant vision

ABSTRACT

Background: From the need for a quantitative method to examine visual processing in young children, we measured ocular orienting responses to visual stimuli (form, motion, expansion, color, contrast, cartoons). Reliability and applicability of this method were assessed.

New method: 80 children (1–12 years) with visual impairments and a reference group of 118 typically developing children (1–6 years) completed two sessions. Test–retest reliability was measured by calculating differences in reaction time and fixation accuracy between the two sessions. For applicability, the cumulative percentage of children that fell within a pre-defined reliability interval was plotted against the mean number of responses.

Results: In typically developing children none of the outcome measures significantly differed between sessions. In the children with visual impairments similar results were obtained, except for motion. This stimulus elicited significantly faster reaction times in the second session. In at least 80% of the children reliable reaction times could be calculated if 4 responses to a cartoon stimulus and 1 or 2 responses to the other stimuli were measured.

Comparison with existing method(s): The existing method to quantify visual information processing has been refined: the range of visual functions was extended and a criterion for reliable assessment of orienting response times was established.

Conclusions: Objective measurement of orienting responses is a reliable method to test the initial stage of visual processing in children with and without visual impairments. A set minimum number of responses for each stimulus warrants the reliability of measurements obtained with this functional method in clinical practice.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Visual information processing is a prerequisite for interaction with and interpretation of the environment. The neural system for

visual information processing involves pathways leading from the retina, through the thalamus, to the striate and extrastriate cortex (Callaway, 2004; Kaplan, 2004). The integrity of this system is crucial: damage to one of the involved areas can cause atypical development of not only visual, but also cognitive and social skills (Braddick and Atkinson, 2011).

With the improvement in perinatal care and higher survival rates in infants, the incidence of cerebral damage and brain development disorders in children have increased. Accordingly, neurological comorbidity in the form of visual processing impairments has become highly prevalent in children and is collectively

* Corresponding author at: Department of Neuroscience, Room Ee1453, Erasmus MC, PO Box 2040, 3000 CA Rotterdam, The Netherlands. Tel.: +31 10 704 3368; fax: +31 10 704 7309.

E-mail addresses: m.kooiker@erasmusmc.nl (M.J.G. Kooiker), j.vandersteen@erasmusmc.nl (J. van der Steen), j.pel@erasmusmc.nl (J.J.M. Pel).

called cerebral visual impairment (CVI). Luckily, children with CVI have the capacity to improve vision over time with age (Khetpal and Donahue, 2007). Given the potential for improvement in such an important area of development, measures of visual processing that are objective, quantitative, valid and reliable are a necessity.

Various methods are currently used to measure vision in infants, e.g. structured observations, visual function tests, visual perception tests and psychophysical methods such as Visual Evoked Potentials (VEP; Good, 2001). Despite their usefulness in clinical practice, each of these methods has its drawbacks. These range from the need for communication and long or subjective assessment to a limited range of visual functions that can be tested. Recently, we showed the feasibility to quantitatively measure the quality of visual information processing with remote eye tracking (Pel et al., 2010a; Boot et al., 2012).

This method comprises measurements of eye movements that are made with a mobile eye tracking set-up. Eye movements are recorded during the presentation of a preferential looking paradigm. Preferential looking is based on the human preference for patterned surfaces to plain ones, and for stimuli which differ from their background (Sturm et al., 2011). When the information of visual stimuli is processed by visual brain areas, orienting eye movements are made automatically toward these stimuli (Boot et al., 2012). The main clinical advantages of this method are that it is quantitative and objective, and that it does not require communication with the child. From the eye movement responses two objective parameters are calculated: reaction time to fixation (RTF) which represents the timing of visual processing, and gaze fixation area (GFA) which represents fixation quality (Pel et al., 2010a).

Previous research showed that children with CVI had slower orienting responses to a variety of visual stimuli than typically developing children and than children with ocular motor impairments (Pel et al., 2010b, 2013). The slower orienting responses could not be explained by other types of visual impairments, i.e. low visual acuity or ocular motor impairments (Kooiker et al., 2014). Thus, on the basis of orienting responses inferences about specific impairments in children with CVI can be made.

Good test–retest reliability of this method has been shown in typically developing children (Pel et al., 2010a; Boot et al., 2012). Both Pel et al. and Boot et al. used the Intraclass Correlation Coefficient (ICC) to quantify reliability (Cicchetti and Sparrow, 1981). Pel et al. (2010a) found good reliability of the minimum RTF and GFA for cartoons, reflected by ICC's of 0.84 and 0.80. They concluded that the eye tracker is well suited for the measurement of timing and execution of oculomotor fixations. Boot et al. (2012) measured test–retest reliability of RTF to form and motion stimuli in typically developing children. ICC's of 0.79 and 0.89 were obtained, indicating good reliability for both form and motion. The authors emphasized that the analyses for motion were based on a low number of observations. In addition, the reliability has not yet been addressed in children with visual impairments.

The aim of the present study was to investigate the test–retest reliability of orienting responses to six stimuli in a large group of children with various types of visual impairments, and to compare these results with the reliability in a reference group of typically developing children. We tested children with visual impairments and typically developing children on two separate occasions. RTF and GFA values in both sessions were compared. We distinguished two subgroups: children with ocular visual impairment (OVI) and children with CVI. In these subgroups we analyzed test–retest reliability using the cartoon stimuli only, because we previously showed delayed responses to cartoons in children with CVI, not to other stimuli (Pel et al., 2010b; Kooiker et al., 2014). In addition, we calculated the minimum number of responses to obtain reliable RTF values for each visual stimulus. We hypothesized that results of both typically developing children and children with

visual impairments are reproducible and reliable over time. We expected test–retest reliability to be better in children with OVI than with CVI.

2. Materials and methods

2.1. Participants

We recruited 119 children with visual impairments between 1 and 12 years at Royal Dutch Visio: an institution for blind and visually impaired people. Visual impairments could originate from CVI, ocular pathology or a combination. Exclusion criteria were a visual acuity of <0.05 (decimal scale), severe physical disabilities and severe epilepsy. A group of 80 children was retested to assess reliability (50 boys, 30 girls, mean age (SD) = 8.6 (2.8)). We refer to all the children with visual impairments as the 'risk group'. Test–retest reliability was analyzed for the total risk group and separately for two subgroups: children with CVI ($n = 22$) and ocular visual impairment (OVI; $n = 37$), according to the diagnosis in the medical records of the children. The remaining 21 children had a combination of visual impairments and were neither classified as CVI nor OVI. We recruited 118 typically developing children (62 boys, 56 girls, mean age (SD) = 3.1 (1.9)) at regular daycare centers. These children served as a reference group. Children in all groups were retested on the same day to assess test–retest reliability. The experimental procedures were approved by the Medical Ethical Committee of the Erasmus University Medical Center, Rotterdam, the Netherlands (MEC-2012-097). The study adhered to the Declaration of Helsinki for research involving human subjects.

2.2. Measurement setup and procedure

The measurement setup consisted of a 24-in. monitor with an integrated infrared eye-tracking system sampling at 60 Hz, to measure visual orienting responses (Tobii T60-XL, Tobii Corporation, Sweden). The system measured gaze position of each eye separately using cornea reflection and compensated free head-movements (system latency ± 30 ms). Each child sat at approximately 60 cm distance from the monitor, to ensure signal reception and efficient tracking of the eyes. Depending on their physical ability the children were seated either in a comfortable chair or in their wheelchair or pram. The experiments were conducted in a quiet room with ambient light conditions. First, a standardized 5-point calibration procedure of both eyes was performed. If this calibration failed the eyes were post-calibrated before starting the analyses. Next, short movies were shown on the eye tracker's monitor in two sessions of both ~ 7 min. All children were shown cartoons, coherent form, motion detection, motion expansion, color and contrast stimuli. Each stimulus had a target area of six degrees, only the form and motion expansion stimuli were slightly larger with a target area of eight degrees. The two sessions consisted of 72 stimulus presentations and contained the same stimuli, only the order in which they appeared was randomized. Time between sessions was on average 15 min. A more detailed description of the stimuli is presented in Appendix. All measurements were stored on hard disk and analyzed off-line using self-written Matlab programs (Mathworks Inc., Natick, MA, USA).

2.3. Data analysis and statistics

The focus of gaze data analysis was to measure reflexive orienting eye movements to the target area of the six stimuli as an estimate of orienting response time and fixation performance. Orienting response time was expressed as the reaction time to fixation (RTF) on a stimulus. For each stimulus

presentation, RTF was defined as the time from the appearance of a stimulus on the monitor ($t=0$), until gaze entered the target area. The gaze fixation area (GFA) was expressed as mean diameter in degrees and used as a measure for fixation accuracy (for a detailed description of the calculation of RTF and GFA we refer to [Appendix](#)).

For the reliability analysis, we used mean RTF and GFA values of the first (RTF1 and GFA1) and the second session (RTF2 and GFA2). Children of which gaze was properly recorded were included for analysis of the specific stimuli. To check for initial differences in the quantity of obtained gaze data, we compared the percentage of gaze data between the two sessions. We measured test–retest reliability in typically developing children, the total group of children with visual impairments and separately for the children with CVI and OVI (only for cartoons). We compared mean differences between sessions with paired samples t -tests, and used an adaptation of Bland–Altman difference plots to investigate systematic bias ([Bland and Altman, 1999](#); [Vaz et al., 2013](#)). Bland–Altman plots traditionally show the mean differences and the 95% limits of agreement between two measurement sessions. These limits represent the interval within which 95% (=1.96 SD) of the mean differences between test sessions are expected to fall in any future sample. We chose to adopt a more stringent approach and used limits of 1 SD to represent the interval within which 68% of the mean differences between sessions are expected to fall. We calculated the limits of the reference group and applied these to both the reference group and the risk group. The lower bound is the mean difference – 1 SD, the upper bound is the mean difference + 1 SD. If a measurement score falls within these limits, it is assumed there is agreement in results between the two test sessions. In addition, we did a univariate analysis of variance to investigate the effect of age on test–retest reliability.

The SD of the differences between sessions was also used to define a reliability interval as a criterion for reliably obtained reaction times of each presented stimulus. First, we plotted the difference between RTF in session 1 and 2 as a function of the mean number of responses in two sessions, for the risk group and reference group together. Second, the mean \pm 1 SD of the mean RTF difference in the reference group was selected as a reliability interval. As a last step, we calculated the cumulative percentage of children whose individual mean difference did not fall outside the reliability interval. We took the corresponding number of responses as a criterion for applicability.

3. Results

3.1. Participants

In the risk group of children with visual impairments, gaze data of 80 children could be included for analysis of cartoons, data of 69 children for contrast, 64 for form, 34 for expansion, 52 for motion, and 41 for color. The percentage of gaze data did not significantly differ between the first and second test session, for any of the stimuli. In the reference group, gaze data of 118 children could be included for cartoons, data of 71 children for contrast, 99 for form, 58 for expansion, 72 for motion and 31 for color. Demographic and clinical factors of children in the risk group and the reference group are presented in [Table 1](#). In both groups the percentage of gaze data was significantly lower in the second session than in the first, for every stimulus. The mean chronological age of children in the reference group was significantly lower than in the risk group (respectively 3.1 and 8.6 years). [Fig. 1](#) shows the relationship between age and the RTF difference score. A summary of the mean values and differences between sessions per stimulus and per group are shown for RTF in [Table 2](#) and for GFA in [Table 3](#). A

Table 1

Demographic and clinical factors of the children with and without visual impairments.

Factor	Risk group	Reference group
Age (mean in years)	8.6 (2.8)	3.1 (1.9)
Gender (m:v)	50:30	62:56
Origin of visual impairments (N)		n/a ^d
CVI ^a	22	
OVI ^b	37	
Combination	21	
Visual acuity (mean, decimal scale)	0.35	– ^e
Nystagmus (N)	46	0
Strabismus (N)	42	0
Intelligence level (mean IQ value ^c)	84	– ^e

^a CVI = cerebral visual impairment.

^b OVI = ocular visual impairment.

^c IQ = intelligence quotient.

^d n/a = not applicable.

^e – = unknown.

visualization of these results is provided with difference plots in [Fig. 2](#) for the cartoon stimulus and the form stimulus.

3.2. Reliability of response time

In the reference group the RTF values did not significantly differ between sessions for any of the stimuli (paired samples t -tests; all $ps > .05$). In the risk group, the RTF values of the stimuli cartoon, contrast, form, expansion and color did not significantly differ between sessions (paired samples t -tests; all $ps > .05$). The RTF values of the motion stimulus were significantly lower in session 2 compared to session 1 ($t = 2.21$, $p < .05$). In the children with CVI and OVI, RTF to cartoons did not significantly differ between sessions (paired samples t -tests, $ps > .05$). [Fig. 2](#) shows that in the reference group, the limits of agreement for RTF to cartoons range from –48 to +46 ms, whereas for the form stimulus the limits are wider, ranging from –434 ms to 318 ms. Data of 52 children in the risk group (65%) fell within these limits for the cartoons, whereas data of 48 children (75%) fell within the reference limits for form. There was no effect of age on test–retest reliability ($F = 0.59$, $p > .05$).

3.3. Reliability of fixation performance

In the reference group as well as the risk group, the GFA values did not significantly differ between the two test sessions, for any of the stimuli (paired samples t -tests, all $ps > .05$).

3.4. Criterion for reliable response times

The variability in response times of all stimuli depended on the average difference between sessions, i.e. variability increased with increasing mean differences. This is shown in [Fig. 3](#) for the cartoons and in [Fig. 4](#) for form. The upper panels of [Figs. 3 and 4](#) illustrate that the average differences in response times tend to decrease with an increasing number of responses, in the risk group as well as the reference group. The lower panels of [Figs. 3 and 4](#) present the successive improvement in reliability of RTF plotted against the increasing mean number of responses. Seen from right to left, every time the RTF difference value of a child fell outside the 1 SD reliability interval, the percentage of children with a reliable RTF value decreased. The corresponding criteria for 80%, 85%, 90% and 95% reliability of RTF for all stimuli are shown in [Table 4](#). A reliable RTF value in at least 80% of the children corresponds with 4 responses to cartoons as a minimum (out of 16). Reliable RTF values for the other stimuli (contrast, color, form, motion and expansion) correspond with 1–2 responses as a minimum (out of 4). When increasing the

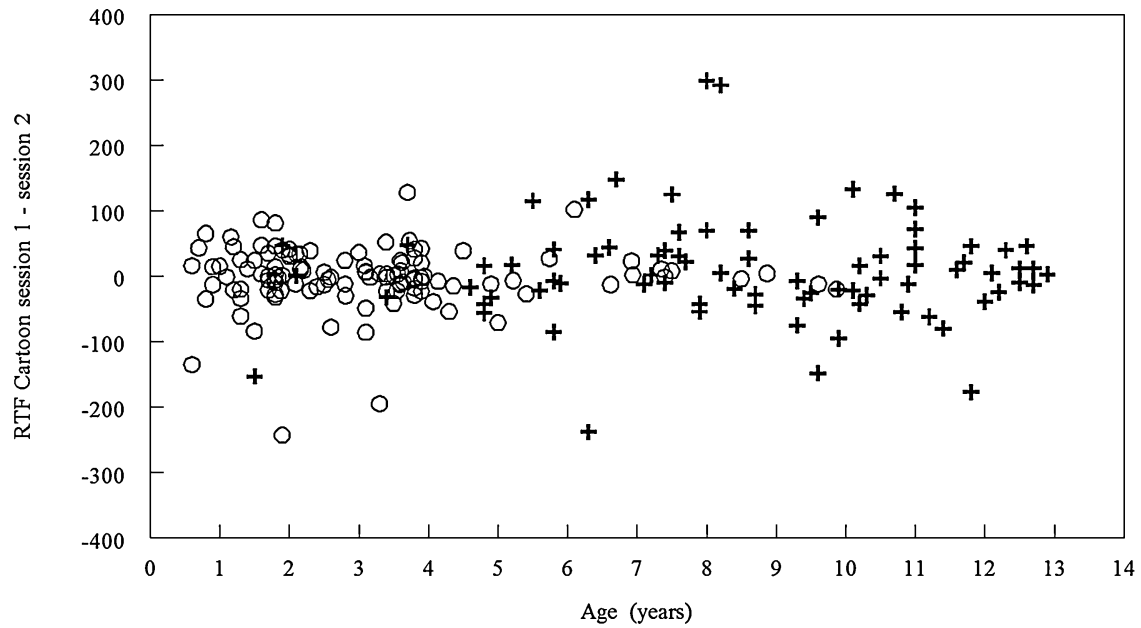


Fig. 1. Mean RTF difference scores of cartoons plotted against age of the children in the risk group (crosses) and reference group (open circles).

Table 2

Mean RTF in session 1 and 2 and the mean differences between sessions for RTF of different visual stimuli in the risk group and reference group.

Stimulus	Group (N)	Mean RTF 1 ^a (SD) ^c	Mean RTF 2 ^b (SD) ^c	Mean RTF difference (SD) ^c
Cartoon	Risk (80)	300 (107)	292 (94)	8 (81)
	Control (118)	223 (41)	224 (52)	-1 (47)
Contrast	Risk (69)	429 (152)	426 (151)	3 (170)
	Control (71)	367 (87)	372 (103)	-5 (119)
Form	Risk (64)	656 (261)	647 (314)	9 (343)
	Control (99)	681 (317)	739 (462)	-58 (376)
Expansion	Risk (34)	718 (289)	762 (536)	-44 (505)
	Control (58)	1105 (670)	1064 (685)	42 (749)
Motion	Risk (52)	928 (511)	773 (406)	155 (506) [*]
	Control (71)	513 (117)	530 (167)	-17 (168)
Color	Risk (41)	955 (422)	883 (404)	72 (295)
	Control (31)	1277 (648)	1121 (568)	156 (740)

^a RTF 1 = reaction time to fixation in the first session.

^b RTF 2 = reaction time to fixation in the second session.

^c SD = standard deviation.

^{*} Significant difference ($p < .05$) between sessions.

Table 3

Mean GFA in session 1 and 2 and the mean differences between sessions for GFA of different visual stimuli in the risk group and reference group.

Stimulus	Group (N)	Mean GFA 1 ^a (SD) ^c	Mean GFA 2 ^b (SD) ^c	Mean GFA difference (SD) ^c
Cartoon	Risk (80)	3.10 (0.92)	3.05 (0.76)	0.05 (0.65)
	Control (118)	1.78 (0.32)	1.84 (0.35)	-0.05 (0.36)
Contrast	Risk (69)	2.58 (0.85)	2.63 (0.98)	-0.05 (0.96)
	Control (71)	1.59 (0.41)	1.59 (0.48)	0.004 (0.63)
Form	Risk (64)	3.79 (1.3)	3.64 (1.3)	0.15 (1.16)
	Control (99)	2.53 (0.82)	2.45 (0.85)	0.074 (1.03)
Expansion	Risk (34)	3.01 (1.19)	2.88 (1.38)	0.13 (1.31)
	Control (58)	1.93 (0.81)	1.97 (0.87)	-0.03 (1.02)
Motion	Risk (52)	2.58 (0.88)	2.56 (0.86)	0.02 (1.11)
	Control (71)	1.59 (0.47)	1.63 (0.48)	-0.05 (0.62)
Color	Risk (44)	2.71 (0.81)	2.86 (1.05)	-0.15 (0.97)
	Control (33)	1.47 (0.75)	1.44 (0.73)	-0.07 (1.05)

^a GFA 1 = gaze fixation area in the first session.

^b GFA 2 = gaze fixation area in the second session.

^c SD = standard deviation.

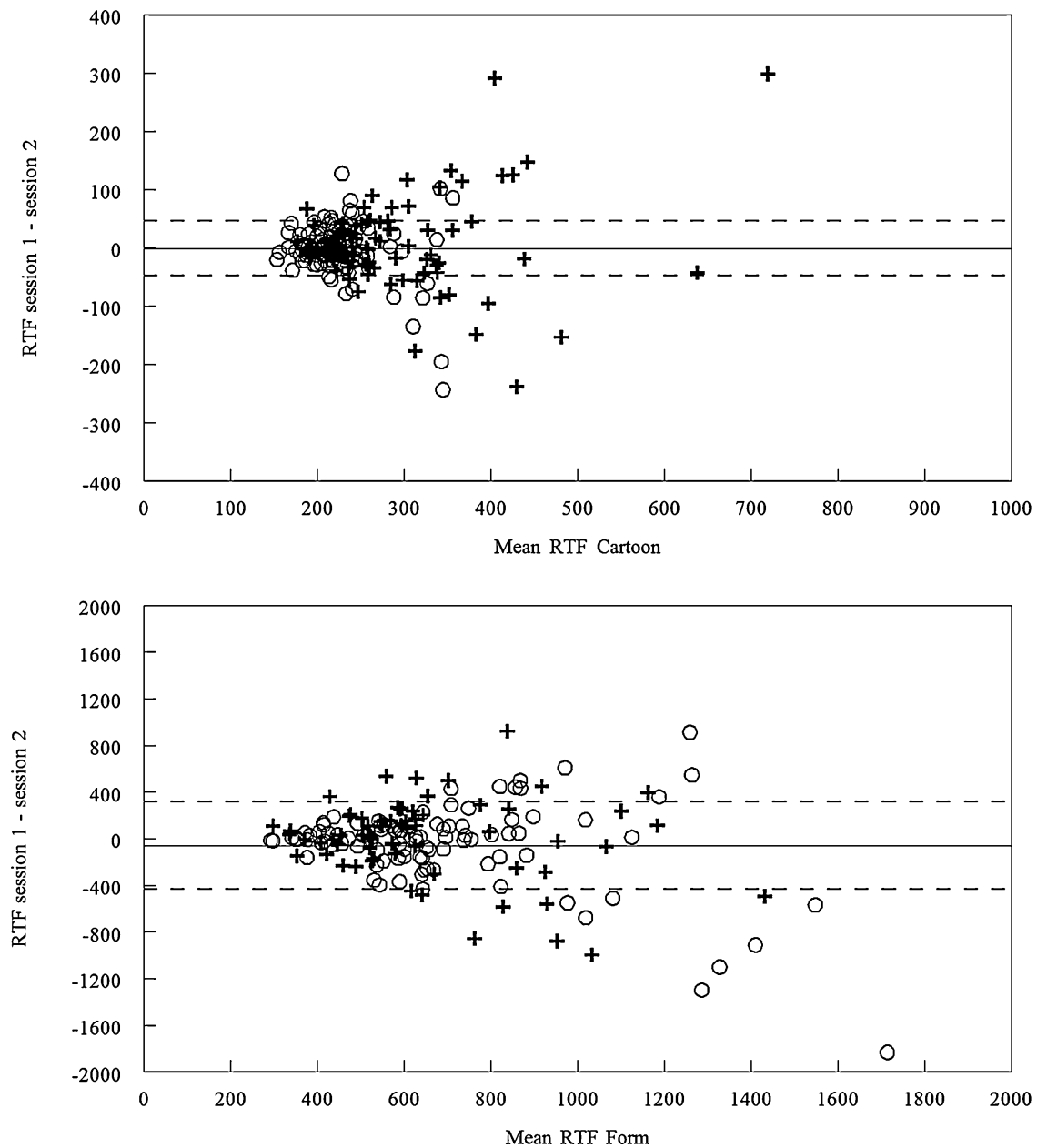


Fig. 2. Bland–Altman plots of the cartoon stimulus (upper panel) and the form stimulus (lower panel), for the risk group (crosses) and reference group (open circles). The plots represent the individual differences in RTF value plotted against the mean RTF value. Dotted lines represent the upper and lower 68% limits of agreement between test sessions. Note the differences in axes values between the upper and lower panel.

Table 4

The number of responses for a reliability of RTF of respectively 80%, 85%, 90% and 95% in the total group of children.

Stimulus	Responses for 80% reliability	Responses for 85% reliability	Responses for 90% reliability	Responses for 95% reliability
Cartoon	4	6	7	8.5
Contrast	1.5	1.5	2.5	3
Form	1.5	1.5	2	2.5
Expansion	1	1	1.5	3
Motion	2	2	2.5	2
Color	1	1	1.5	1.5

reliability to 85% only the criterion for cartoons increases, whereas at a reliability of 90% criteria for all stimuli increase.

4. Discussion

We examined the test–retest reliability of orienting response times (RTF) and fixation accuracy (GFA), two outcome

parameters provided by the eye tracking method, in children with visual impairments and a group of typically developing children. RTF was stable over time for most stimuli, while GFA was stable over time for all stimuli. RTF measures were reliable in children with OVI as well as CVI. Criteria to obtain reliable data provide a foundation for the feasibility of the method in clinical practice. These findings not only have implications for the method under

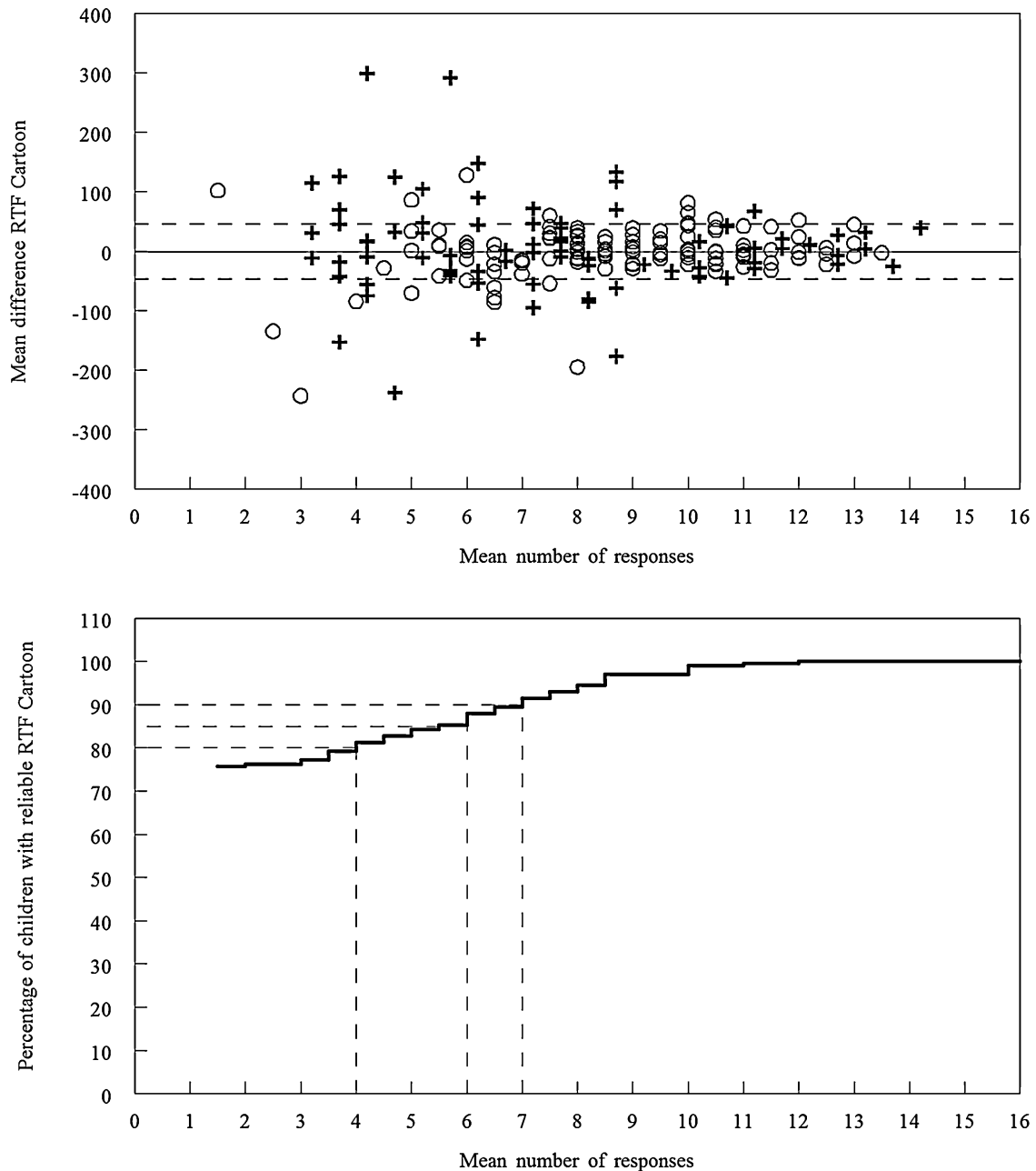


Fig. 3. Upper panel: mean RTF difference between sessions in the risk group (dots) and reference group (crosses) plotted against the mean number of responses to cartoons. Horizontal lines represent the reference reliability interval, i.e. the mean difference ± 1 SD of the reference group. Lower panel: the cumulative percentage of children with reliable data for cartoons (mean difference within the reliability interval of the upper panel) plotted against the mean number of responses in two sessions for the total group of children.

study, but also for the stability of functioning over time in children with visual impairments.

4.1. Reliability and criterion for response times

The outcome parameters in the current study are reflexive ocular motor responses to stimuli, so called bottom-up responses. In humans, a minimum visual processing time of about 100 ms is needed before an eye movement is executed (Fischer and Ramsperger, 1984).

In the risk group, response times to the motion stimulus were significantly faster during the second test session than in the first. Such a decrease in reaction times is generally explained by a learning effect. The target area of each stimulus was randomly selected

which disabled the possibility to predict the stimulus locations. Therefore, the learning effect cannot be explained in terms of predictability. A possible explanation might be that the target of this stimulus is relatively difficult to discover when presented for the first time. In the reference group we do not see a decrease in response time for motion. This stimulus might be more difficult to detect for children with visual (processing) impairments, who have slower response times and larger fixation areas for this stimulus than typically developing children, in this study and a previous one (Pel et al., 2013). These difficulties should be taken into account in future applications of this method, for example by providing a practice session before the start of the measurement.

In contrast to motion, response times of the cartoons, contrast, expansion, color and form stimuli did not change over

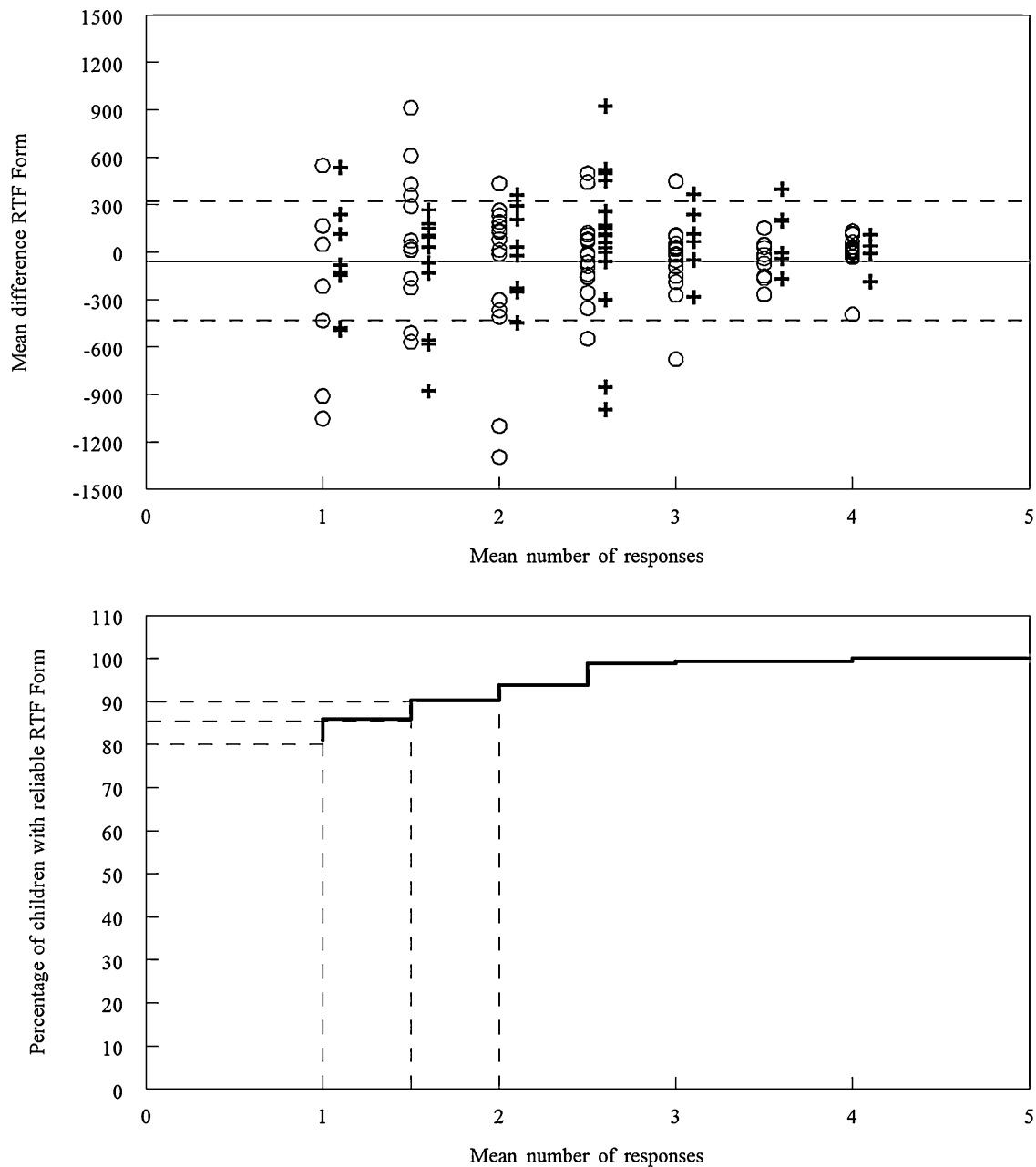


Fig. 4. Upper panel: mean RTF difference between sessions in the risk group (dots) and reference group (crosses) plotted against the mean number of responses to form. Horizontal lines represent the reference reliability interval, i.e. the mean difference ± 1 SD of the reference group. Lower panel: the cumulative percentage of children with reliable data for form (mean difference within the reliability interval of the upper panel) plotted against the mean number of responses in two sessions for the total group of children.

measurement sessions. A practical consequence of this result is that these stimuli can be presented less often in each test session than currently done. It reduces the 14 min testing time of the current method by a factor 2, which improves the possibility to successfully test children with a short attention span.

One approach for applicability of this test would have been the definition of a cut-off point related to a maximum reaction time and the corresponding degree of variability. For example, when looking at the difference plot of cartoons in Fig. 2, a cut-off value of 300 ms could have been a sensible threshold between reliable and unreliable measurements, since most values that exceed 300 ms show high variability. However, certain visual impairments are associated with delays in RTF (for example nystagmus and CVI; Pel et al., 2010b, 2013). Hence, the RTF seems to reflect true differences between subjects and groups of patients and can therefore

not be used to infer a criterion for reliable data. Instead, we chose a performance measure to quantify applicability, i.e. the mean number of responses as a function of corresponding mean differences in RTF over time. In Figs. 3 and 4 it is shown that the mean RTF differences as a function of responses show a similar pattern in the control and risk group. For cartoons, 4 responses (out of 16) are needed to warrant good reliability in at least 80% of the children in the tested group. To achieve similar reliability for the other stimuli, 1–2 responses (out of 4) are needed. This corresponds with 25–50% of the presented stimuli. Although in the current study we showed cartoons 16 times per session to enable post calibration and enhance attention, reliability would not be compromised when they are shown less often. The reference reliability intervals to establish these criteria were derived from the reference group, which consists mainly of young children (below the age of 6). We

know from previous studies that response times to form and motion decrease until the age of 6–8, after which they reach mature levels (Boot et al., 2012). Therefore, we expect the reliability criteria to become more stringent with increasing age. The currently reported criteria in the reference group are valid for young children and they merely provide a lower threshold of reliability for older children. When improving the degree of reliability to 90%, the criteria also become more stringent. Where to put the reliability criterion is in essence a clinical question, depending on the group under study. In the current risk group, with a broad range and severity of visual and developmental impairments, we believe that reliable scores in 80% of the children suffice.

Fig. 2 shows that the 68% limits of agreement for RTF in the reference group ranged from about 50 ms for cartoon to 400 ms for form. The discrepancy between the two sessions (the wideness of the limits) is small enough to be clinically important for most stimuli: it enables stable and reproducible measurements of visual responses. The present results, and the fact that our method is sensitive enough to pick up group differences, provide a foundation for use and implementation in clinical practice.

4.2. Reliability of response time in OVI vs. CVI

In children with OVI and CVI mean response times did not significantly differ between test sessions. This was to be expected for the children with OVI, who have in general relatively stable eye conditions. In contrast, in CVI there is a large body of literature that emphasizes its fluctuating nature, reporting that children with CVI have vision that seems to change over time, even within one day (Dutton, 2003; Dutton et al., 2004, 2006; Good and Hoyt, 1989; Jan et al., 1993; Matsuba and Jan, 2006). We assessed both sessions on the same day at the rehabilitation centers (risk group) or at daycare institutions (reference group) to keep interference with the children's daily structure to a minimum. This was probably an advantage in the case of children with CVI, given its reported fluctuating nature. We previously showed prolonged response times in children with CVI, which points toward a compensational strategy for a lack of parallel processing capacity (Kooiker et al., 2014). The current findings on test–retest reliability of this method indicate that it might be a stable, and thus promising, measure of the initial stage of visual processing in CVI. Future research will be directed toward the visual development of orienting responses over years instead of days, and toward later stages of visual processing.

4.3. Reliability of fixation performance

In the reference group as well as the risk group, the parameter fixation accuracy was a reliable measure over time for all stimuli we used. This warrants the prediction of fixation performance in future samples of children. Fixation performance is a fundamentally different measure compared to RTF, and might reflect a more inherent ability which is not susceptible to fluctuations over time. An example is the ocular motor disorder (congenital) nystagmus, in which less accurate fixations are probably inherent to the dysfunction. Various studies have reported fixation instabilities in children with nystagmus (Aring et al., 2007; Dell'Osso et al., 1992; Huurneman and Boonstra, 2013). Accordingly, we repeatedly showed a relationship between decreased fixation quality and the presence of nystagmus in children (Pel et al., 2010b, 2013).

4.4. Study limitations

An important issue is the age difference between the groups: typically developing children were younger than children in the risk group. In typically developing children over the age of 4 years, reaction times are expected to be faster and show less variability

within and between subjects than in younger children. This has been shown previously for expansion and form (Boot et al., 2012) and cartoons (Pel et al., 2010a). The age difference between groups explains why the limits of agreement were relatively large in the reference group, and why the mean difference scores do not differ greatly from those in the risk group. However, the reliability of response times was not affected by age in our population. The method we used is directed toward the early detection and identification of visual information processing impairments. In the reference group it was relatively easy to include young children. In the risk group less young children were available to be recruited. In current visual rehabilitation practice, visual processing impairments are often not detected until the age of 5 or 6. By providing reference criteria based on young typically developing children we take an advance in the clinical implementation of the method, with the ultimate goal of early detection of visual information processing impairments in young children.

5. Conclusion

The quantitative measure of orienting responses is a reliable method to test the initial stage of visual processing in typically developing children and children with a variety of visual impairments. The results of test–retest reliability and the proposed criterion for reliable data interpretation provide a foundation for application and implementation of the method in clinical practice.

Financial support

Financial support for this study was provided by 'ZonMw Inzicht' (Netherlands Organization for Health Research and Development – Insight Society), grant number: 60-00635-98-10.

Acknowledgements

The authors would like to thank the children and their parents, and all involved people from Royal Dutch Visio and the daycare centers (Wasko, Alblasterwaard) for their cooperation. We thank our colleague Mark Vonk for his help in data collection.

Appendix

Stimuli

All stimuli contained a specific area in one of the quadrants of the monitor, defined as the target area. Cartoons were shown 16 times per session, because they were used for post-calibration of the data, whereas all other stimuli were shown 4 times per session. Each stimulus was shown for 4 s and the position of the target area was varied at random, in such a way that all stimuli appeared at least once in each quadrant of the monitor.

For a detailed description of the stimuli that were presented we refer to previous studies (Pel et al., 2010a,b, 2013; Kooiker et al., 2014).

Outcome parameters

Response time was expressed as the reaction time to fixation on a stimulus.

If gaze was already in the target area at $t = 0$, these eye movements were excluded from further analyses. A cumulative plot was made from the available reaction time values per stimulus. An exponential curve was fitted to this cumulative plot to quantify the reaction time as the minimum reaction time plus 1/3 of the time constant of the exponential curve. This value was denoted as the

reaction time to fixation (RTF). Fixation performance was defined as 85% of all gaze coordinates within the target areas using principal component analysis (Oliveira et al., 1996).

References

- Aring E, Grönlund MA, Hellström A, Ygge J. Visual fixation development in children. *Graef Arch Clin Exp* 2007;245:1659–65.
- Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999;8:135–60.
- Braddick O, Atkinson J. Development of human visual function. *Vis Res* 2011;51:1588–609.
- Boot FH, Pel JJM, Evenhuis HM, van der Steen J. Quantification of visual orienting responses to coherent form and motion in typically developing children aged 0–12 years. *Invest Ophthalmol Vis Sci* 2012;53:2708–14.
- Callaway EM. Cell types and local circuits in primary visual cortex of the macaque monkey. In: Chalupa LM, Werner JS, editors. *The visual neurosciences*. Cambridge, MA: MIT Press; 2004. p. 680–94.
- Cicchetti DV, Sparrow SA. Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *Am J Ment Defic* 1981;86:127–37.
- Dell'Osso LF, van der Steen J, Steinman RM, Collewijn H. Foveation dynamics in congenital nystagmus. I: Fixation. *Doc Ophthalmol* 1992;79:1–23.
- Dutton GN. Cognitive vision, its disorders and differential diagnosis in adults and children: knowing where and what things are. *Eye* 2003;17:289–304.
- Dutton GN, Saaed A, Fahad B, Fraser R, McDaid G, McDade J, et al. The association of binocular lower visual field impairment, impaired simultaneous perception, disordered visually guided motion and inaccurate saccades in children with cerebral visual dysfunction—a retrospective observational study. *Eye* 2004;18:27–34.
- Dutton GN, McKillop EC, Saidkasimova S. Visual problems as a result of brain damage in children. *Br J Ophthalmol* 2006;90:932–3.
- Fischer B, Ramsperger E. Human express saccades: extremely short reaction times of goal directed eye movements. *Exp Brain Res* 1984;57:191–5.
- Good WV, Hoyt CS. Behavioural correlates of poor vision in children. *Int Ophthalmol Clin* 1989;29:58–60.
- Good WV. Development of a quantitative method to measure vision in children with chronic cortical visual impairment. *Trans Am Ophthalmol Soc* 2001;99:253–69.
- Huurneman B, Boonstra FN. Monocular and binocular development in children with albinism, infantile nystagmus syndrome and normal vision. *Strabismus* 2013;21:216–24.
- Jan JE, Groenvelde M, Anderson DP. Photophobia and cortical visual impairment. *Dev Med Child Neurol* 1993;35:473–7.
- Kaplan E. The M, P, and K pathways of the primate visual system. In: Chalupa LM, Werner JS, editors. *The visual neurosciences*. Cambridge, MA: MIT Press; 2004. p. 481–93.
- Khetpal V, Donahue SP. Cortical visual impairment: etiology, associated findings, and prognosis in a tertiary care setting. *J AAPOS* 2007;11:235–9.
- Kooiker MJG, Pel JJM, van der Steen J. Viewing behavior and related clinical characteristics in a population of children with visual impairments in the Netherlands. *Res Dev Disabil* 2014;35:1391–401.
- Matsuba CA, Jan JE. Long-term outcome of children with cortical visual impairment. *Dev Med Child Neurol* 2006;48:508–12.
- Oliveira LF, Simpson DM, Nadal J. Calculation of area of stabilometric signals using principal component analysis. *Physiol Meas* 1996;17:305–12.
- Pel JJ, Manders JC, van der Steen J. Assessment of visual orienting behaviour in young children using remote eye tracking: methodology and reliability. *J Neurosci Methods* 2010a;189:252–6.
- Pel JJM, Does LV, Boot F, Faber TD, van der Steen-Kant SP, Willemsen S, et al. Effects of visual processing and congenital nystagmus on visually guided ocular motor behaviour. *Dev Med Child Neurol* 2010b;53:344–9.
- Pel JJ, Kooiker MJ, Does JM, Boot FH, Faber JT, van der Steen-Kant SP, et al. Orienting responses to various visual stimuli in children with visual processing impairments or infantile nystagmus syndrome. *J Child Neurol* 2013. <http://dx.doi.org/10.1177/0883073813511150>.
- Sturm V, Cassel D, Eizenman M. Objective estimation of visual acuity with preferential looking. *Invest Ophthalmol Vis Sci* 2011;52:708–13.
- Vaz S, Falkmer T, Passmore AE, Parsons R, Andreou P. The case for using the repeatability coefficient when calculating test-retest reliability. *PLoS ONE* 2013;8:e73990.