Published in final edited form as:

J Vestib Res. 2010 ; 20(5): 363–372. doi:10.3233/VES-2010-0371.

The reliability, stability, and concurrent validity of a test of gaze stabilization

Bryan K. Ward^a, Maha T. Mohammad^b, Susan L. Whitney^{a,b}, Gregory F. Marchetti^{a,c}, and Joseph M. Furman^{a,b}

^aDepartment of Otolaryngology, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

^bDepartment of Physical Therapy, University of Pittsburgh, Pittsburgh, PA, USA

^cRangos School of Health Sciences, Duquesne University, Pittsburgh, PA, USA

Abstract

The gaze stabilization test (GST) is a computerized test of the vestibuloocular reflex that reports maximum head velocity while maintaining fixed visual acuity. The GST thus assesses the vestibuloocular reflex differently from the dynamic visual acuity test (DVAT). The purpose of this study was to assess the reliability, stability, and validity of the GST in a healthy young and older population. Forty subjects (20 older adults with mean (SD) age of 76.3 (5.3) and 20 young controls with mean (SD) age of 25.2 (3.2)) performed the GST and DVAT assessments. The version of the GST used in this study has a tunneled mirror system to ensure a consistent participant distance of 4 m from the computer screen. All subjects repeated trials within 30 minutes of initial testing. Twenty subjects (10 from each age group) returned 7–10 days later to repeat the GST and DVAT assessments. Vestibular symptoms were assessed before and after GST and DVAT assessments. The mean (SD) GST scores for the older group were 123 (33) deg/s in the yaw plane and 108 (27) deg/s in the pitch plane. For the young group, mean (SD) GST scores were 157 (34) deg/s in the yaw plane and 141 (25) deg/s in the pitch plane. There was a significant between-group difference for GST scores in both yaw and pitch planes ($p < 0.01$). The intraclass correlation coefficient (ICC) for GST scores performed on the same day was 0.75 in the yaw plane and 0.69 in the pitch plane. The ICC including the 20 subjects who repeated the GST within 7–10 days was 0.59 in the yaw plane and 0.54 in the pitch plane. In general, GST was more stable than DVAT. GST was more stable in younger vs. older subjects whereas DVAT was more stable in older vs. younger subjects. Concurrent validity, determined by Spearman correlation coefficients between GST and DVAT loss results were -0.62 in the yaw plane and -0.38 in the pitch plane ($p < 0.02$). These results suggest that the gaze stabilization test (GST) has good sameday test-retest reliability and stability in healthy young and older adults. The moderate correlation between sameday GST and DVAT loss scores suggest the two tests may be measuring similar, but different constructs.

Keywords

Gaze stabilization test; vestibuloocular reflex; aging

Introduction

People with vestibular disease report visual disturbances as well as dizziness. One of the more common complaints of these patients is gaze instability while walking or head movement induced dizziness and visual blurring [15,24]. The probable cause of this instability is the inability of the vestibular ocular reflex (VOR) to maintain a steady gaze, resulting in increased retinal slip [15,24,28]. Decrement in visual acuity during head movement is potentially a serious problem, as older adults often become unstable with head movement during gait [29] and persons with vestibular disease fall frequently, regardless of age [16,29].

It is currently impossible to replicate the exact situations that trigger head movement induced dizziness or visual blurring. However, VOR function can be inferred from dynamic visual acuity, i.e., the acuity during head movement [6,16,24,28]. The dynamic visual acuity test (DVAT) is an established clinical measure of vestibular function and is used in this study for concurrent criterion validity assessed against the gaze stabilization test [1,6,16,26]. Prior studies of the computerized DVAT indicate excellent test retest reliability and sensitivity and specificity for identifying unilateral vestibular disease [16]. An additional test of VOR function, the gaze stabilization test (GST), was developed as an alternative to the DVAT.

The GST is a measure of the maximum voluntary rotational head velocity, in units of degrees per second, that allows accurate perception of a visual stimulus. The original formulation of the GST has been shown to be highly reliable and to correlate with DVAT results in a population of patients with unilateral vestibular disease [11,22]. Degrees per second of head movement is a measure that is understood functionally and potentially provides a method of tracking patient improvement over time. To further standardize the GST, adjustments have been made using mirrors to fix the viewing distance at 4 meters to better match traditional visual assessments. This newer version of the GST has no published data in either older or young adults.

To accurately assess the function of the VOR in order to manage patients with vestibular disease, a clinical measure must be reliable, stable over time, and consistent with current clinical standards. The primary purpose of this study was to describe the test retest reliability and within subject stability of GST measurements taken within single testing sessions and between testing sessions 7–10 days apart, as well the concurrent validity of GST with the DVAT. A secondary purpose was to evaluate the effect of age on reliability, stability, and levels of performance for measures of GST taken from younger as compared to older subjects.

Methods and materials

2.1. Subjects

Forty subjects completed this study that had been approved by the University of Pittsburgh IRB. Subjects included 20 young adults (10 males) with mean (SD) age 25.2 (3.2) years, and 20 older adults (10 males) of age 76.3 (5.3) years. Subjects were excluded if they had a history of neurological or otological disease, binocular visual acuity with corrective lenses worse than 20/40 in both eyes, a significant directional preponderance on earth vertical axis rotational testing, or performance worse than 11/2 SD below age adjusted means on 2 or more categories of the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) [23]. Older subjects were recruited from a registry of older adults maintained by the University of Pittsburgh Pepper Center and via advertisements. Young subjects were recruited via advertisements. Fifty five subjects were called to recruit 41 subjects for participation. One subject was unable to complete testing due to technical problems and was unable to reschedule testing. Subjects received payment for participation in this study.

2.2. Power analysis

Using GST data from a prior study by Whitney et al. and assuming that repeatability of the test has an Intra class Correlation (ICC) of 0.75, we determined that a suitable sample size for assessing reliability consisted of approximately 40 subjects [30]. Additionally, at a power level of 0.85, we could detect a difference between the older and younger groups of 70 degrees per second for a sample size of 20 subjects per group.

2.3. Protocol

All participants ($n = 40$) performed the GST and DVAT twice during their first visit. The order of testing was GST pitch, GST yaw, DVAT pitch, then DVAT yaw. Participants were offered breaks between each trial and took at least a five minute break before repeating all tasks on the same day. Half of the participants ($n = 20$, 10 young adults and 10 older adults) returned in 7–10 days to repeat GST and DVAT assessments. Testing was performed using the In Vision system (Neurocom, Portland, OR) that includes assessment of static visual acuity and minimal perception time in addition to the GST and DVAT. Static visual acuity and perception time testing is required for establishing parameters for GST and DVAT assessments and was performed on both test days. During testing, participants were seated upright. Participants focused on a computer screen set at a fixed distance of 4 meters via layered mirrors. Environmental lighting during testing was controlled at approximately 6–7 foot candles or 70 lx (determined by a luxmeter) using an intensity adjustable halogen lamp located 1 meter behind the subject's left shoulder. Assessments of symptom presence and severity including 'dizziness', 'nausea', 'visual blurring', and 'headache' were performed on day 1 using a visual analog scale (VAS). Subjects indicated the presence of each symptom by marking a 10 cm line anchored by "no symptom at all" at zero and "as bad as it can be" at 10 cm before all tests and after GST and DVAT assessments in the yaw and pitch planes. The In Vision™ (Neurocom) software allowed assessment of the number of trial failures caused either by a participant's inability to achieve the minimum required head velocity or their inability to maintain the head velocity throughout the trial. These were recorded as

failures and counted as an incorrect response according by the parameter estimation (PEST) convergence algorithm in the software.

2.4. Static visual acuity

Head fixed static visual acuity (SVA) is required in order to assess changes in visual acuity during head movements. Visual acuity was assessed using the In Vision system at the beginning of each visit. Participants were encouraged to wear corrective lenses if needed during all testing. For each trial, a box was presented for 2 s in the center of the computer screen. The subject directed their gaze to the box, which would disappear and then, after a 200 ms delay, an 'E' optotype would appear for 2 s in the previous location of the box. The subject reported aloud the perceived orientation and their responses were entered into the computer by a technician. A PEST algorithm that adapts the presented optotype size based on the success or failure of the previous trial was used to determine the participant's SVA. The PEST algorithm converges on the smallest accurately identified optotype size [27]. Static visual acuity output is reported in logMAR with 0.0 logMAR approximating 20/20 vision and 0.2 logMAR approximating 20/30 vision. The visual acuity for the GST was set at 0.2 logMar (approximately 2 lines on a visual acuity chart) above the participant's baseline SVA.

2.5. Perception time test

The perception time test (PTT) is the minimum amount of time that an optotype 'E' of a determined size can be presented to a participant such that they can correctly identify its orientation. Participants directed their gaze to a box that would appear in the center of the computer screen for 2 s. The box disappeared and after 200 ms, an 'E' optotype 0.2 logMAR above the subject's SVA appeared in its place for a variable amount of time. Similar to SVA testing, a PEST algorithm was used to vary the dwell time of the 'E' optotype. All testing began with a dwell time of 250 ms and varied in increments of 10 ms. The shortest dwell time at which the optotype was correctly identified was the participant's perception time. The perception time was used to determine the optotype dwell time during the DVAT and GST to ensure that the results of these assessments were not influenced by eye movements other than the VOR. Subjects were excluded if their PTT exceeded 80 ms as this would lead to dwell times that would allow the subject to make a corrective saccade [17]. No subjects were excluded on this basis.

2.6. Gaze stabilization test

The GST was designed to assess how quickly a participant could move their head while maintaining focus on a computer generated target of fixed size. Participants wore a headband with a 3axis integrating gyro (InterSense Inertia Cube2, Engineering Systems Technology, Kaiserslautern, Germany) that determined rotational head velocity in the pitch and yaw planes. Participants generated repetitive head movements in the pitch and yaw plane while maintaining gaze on the center of a computer screen. Two feedback bars that appeared on the screen before each task provided information to the participant and the technician on head velocity and amplitude. Participants had 8 s from the start of head movement to achieve the minimum required head velocity for each trial. When the subject's head velocity exceeded the required minimum threshold for a trial, the feedback bars disappeared and the

optotype was displayed until either the maximal dwell time for the optotype was achieved or the subject's head velocity fell below the velocity requirement for that trial. The maximal dwell time for this optotype was set at 35 ms greater than the subject's perception time if the PTT was at least 40 ms. If the PTT was less than 40 ms, the maximal dwell time was set at 75 ms. The subject attempted to identify the orientation of the 'E'. Responses were recorded by a technician. Participants completed a series of trials in which the minimum required head velocity varied according to a PEST algorithm. If the participant was unable to achieve the minimum required head velocity within 8 s from the start of head movement, if the participant was unable to maintain the minimum head velocity for the required duration, or if the participant achieved and maintained the required velocity but incorrectly identified the orientation, the trial was recorded as a failure by the PEST algorithm. The outcome recorded for the GST was the mean of the three fastest head velocities at which the person could move their head and still accurately identify the orientation of the optotype.

2.7. Dynamic visual acuity test

The computerized DVAT assesses the ability of the VOR to maintain a participant's visual acuity while moving their head with a fixed head velocity requirement. Like the GST, this test also required the use of a headband with a 3axis integrating gyro. Participants generated rotational head movements in the pitch or the yaw plane to at least 20 degrees from midline in each direction. Two feedback bars provided information to the participant and the technician on head turning velocity and amplitude. Once the participant's head velocity achieved a minimum of 60 degrees per second in the pitch plane or 85 degrees per second in the yaw plane, the feedback bars disappeared and an optotype "E" appeared in the center of the screen. The dwell time for the optotype was determined using the same guidelines used for the GST assessment. Participants recited the orientation of the "E", which was then entered by a technician. A PEST algorithm was also used to converge on the smallest accurately identified optotype orientation or minimum visual acuity during head rotation. The results of the DVAT were recorded as the size of the smallest optotype that the participant was able to correctly identify while rotating the head faster than the minimum velocity. DVAT scores were converted to visual loss by subtracting baseline static visual acuity for each visit.

2.8. Statistics

All statistics were performed using STATA (version 9, STATA corp., College Station, TX) and SPSS (version 16, SPSS Inc., Chicago, IL). For purposes of analysis, GST and DVAT performance measurements were reduced to summary pitch and yaw plane directions by averaging the up down and left right values respectively for each testing session. To determine the test retest reliability of the GST and DVAT measurements, intra class correlation coefficients (ICC, two way mixed effects model) with 95% confidence intervals (95% CI) were calculated for two measurements taken from all subjects on day 1. This ICC is a ratio of the between subjects variance from test 1 to test 2 to the total variance (between subjects plus within subjects) from test 1 to test 2. An ICC was also estimated for all GST and DVAT measurements taken from 20 subjects tested on day 1 and day 2. The ICC reflects the relative reliability, or the amount of measurement error within a group of subjects. These correlations between related measurements are interpreted using the following criteria: ICC

> 0.75 is considered “excellent” reliability, $0.4 < ICC < 0.74$ is considered “fair to good reliability” [7]. The effect of age on the difference between repeated testing of DVAT and GST was described by estimating the test retest reliability (ICC) separately for age groups consisting of subjects younger than 60 years and 60 years and older. The stability of DVAT and GST measures across measurement intervals (within day and between days) was estimated using the standard error of the measurement (SEM). The SEM is a measure of the within subject variability (error influence on individual scores) due to repeated measures. The SEM is calculated from the ICC using the formula: $SEM = SD \times \sqrt{1 - ICC}$, where SD=sample standard deviation, and ICC=intra class correlation coefficient and $SEM\% = SEM / \text{mean}$. The SEM was then used to estimate the minimal detectable change (MDC), which is the smallest amount of change that is probably not due to chance variation in measurement, where the $MDC_{95} = SEM \times 1.96 \sqrt{2}$. The value of $MDC \times 95$ describes the amount of true change in subject status beyond measurement error with 95% certainty [14]. Concurrent validity was estimated by describing the strength of the association between participants’ performances on the GST and DVAT using the Spearman rank order correlation coefficient. Mean differences between pitch and yaw plane GST and mean differences between pitch and yaw plane DVAT performances within subjects were compared using a ttest for paired samples. Age group differences in mean GST and DVAT performance were compared using ttest for independent samples. The number of test failures was compared between the pitch and yaw planes using the nonparametric Wilcoxon rank sum test for dependent samples and between age groups using the nonparametric Mann Whitney U test. Median number of subjects reporting pretest symptoms was compared between age groups using a nonparametric sign test.

3. Results

Mean (SD) perception time for all participants was 36.4 (25.7) ms with no difference between the younger and older groups. The mean (SD) static visual acuity for the older group was 0.16 (0.1) logMAR and for the younger group was -0.11 (0.1) logMAR ($p < 0.01$). Table 1 shows mean (SD) GST maximum head velocity and DVAT visual loss scores for both age groups. Mean (SD) GST scores for the younger group (averaged over two tests on day 1) were 157 (34) degrees per second in the yaw plane and 141 (25) degrees per second in the pitch plane. For the older group mean (SD) scores in the yaw plane were 123 (33) degrees per second and in the pitch plane were 108 (27) degrees per second. For subjects overall, mean GST velocity was significantly faster in the yaw plane than in the pitch plane for all tests ($p < 0.01$). For subjects overall, there were no differences between mean pitch and yaw plane measures of DVAT for any test. Mean GST velocity was faster in the yaw plane than in the pitch plane for all tests in younger subjects ($p < 0.03$). Mean GST velocity was faster in yaw plane than in pitch plane in older subjects ($p < 0.02$) for all tests except for day 1 test 1. Mean DVAT measures did not differ between the pitch and yaw planes in either age group. Mean GST velocity recorded in both pitch and yaw planes was significantly faster in the younger subjects compared with older for all tests ($p < 0.03$). There were no differences between age groups in mean visual acuity loss for the DVAT in both pitch and yaw planes.

3.1. Test retest reliability

For subjects overall, test-retest reliability estimates for two measures of the GST within a single session were good to excellent; ICC (95% CI) = 0.69 (0.48–0.82) and 0.75 (0.58–0.86) for pitch and yaw plane, respectively. Test retest reliability estimates for two measures of DVAT within a single session were good; ICC (95% CI) = 0.60 (0.36–0.77) and 0.56 (0.30–0.74) in the pitch and yaw plane, respectively. Test retest reliability estimates for all subjects within a single session were similar between GST and DVAT (Table 2). Both the GST and DVAT reliability estimates within session were similar for younger vs. older age groups (Table 2). Test retest reliability estimates for two measures of the GST from 20 subjects across a 7–10 day interval were fair to good; ICC (95% CI) = 0.54 (0.14–0.79) and 0.59 (0.21–0.81) in the pitch and yaw plane, respectively. Test retest reliability estimates for two measures of the DVAT across a 7–10 day interval were poor fair; ICC (95% CI) = 0.10 (not significant) and 0.49 (0.08–0.76) in the pitch and yaw plane, respectively (Table 2). When evaluated within each age group ($n = 10$ in each group) separately, the ICC estimates for GST and DVAT between sessions in both age groups failed to reach significance at $p < 0.05$ in a test against a true ICC value of zero. (Table 2).

3.2. Test stability

Averaged across all subjects, GST displayed less influence due to measurement error (SEM% 13–15%) compared to DVAT (SEM% 27–53%) (Table 3A). The amount of change that would not be attributable to measurement error, i.e., MDC95, for GST ranged from 51 deg/sec to 55 degrees/sec. The MDC95 for DVAT ranged from 0.12 logMAR to 0.24 logMAR. For younger subjects, GST demonstrated less influence due to measurement error (SEM% 11–15%) than DVAT (SEM% 39–55%). Values of MDC95 for GST velocity ranged from 44 deg/sec to 63 deg/sec. Values of MDC95 for DVAT ranged from 0.16 logMAR to 0.21 logMAR (Table 3B). For older subjects, GST demonstrated less influence due to measurement error (SEM% 13–20%) than DVAT (SEM% 31–53%). Values of MDC95 for GST velocity ranged from 44 deg/sec to 64 deg/sec. Values of MDC95 for DVAT ranged from 0.15 logMAR to 0.25 logMAR (Table 3C).

3.3. Concurrent validity

Concurrent validity as measured by Spearman correlation coefficients between GST scores and visual loss in both the pitch and yaw plane are shown in Fig. 1. Overall, DVAT visual loss and GST scores were moderately inversely correlated in pitch and yaw planes with an r value of -0.38 and -0.62 respectively ($p < 0.02$). When separated by age group, significant moderate correlations were detected between DVAT visual loss and GST scores in the yaw plane for older adults ($r = -0.74$, $p < 0.01$) and for younger adults ($r = -0.46$, $p < 0.05$). In the pitch plane there was no significant association between GST scores and measures of DVAT visual loss in either age group.

3.4. Trial failures

For all subjects, the median number of GST failures was greater in the pitch plane (3–4 test failures) as compared to the yaw plane (1–2 test failures) across all three tests. This difference in the number of GST failures was significant for test 1 and 2 ($p < 0.01$) but not

for test 3. The median number of failures was greater in older subjects than in younger subjects in both the pitch and the yaw plane for all three tests ($p < 0.01$).

3.5. Symptoms

No subject reported nausea at any time. Only one subject reported visual blurring (3 cm) on VAS before testing. Two subjects reported mild headache during testing that resolved by test completion. Dizziness was the most prevalent symptom, with 13 subjects (33%) reporting dizziness at some point during the testing session. No more than 20% of the subjects reported dizziness following any particular test. The median reported dizziness level did not increase significantly from baseline to posttest for the entire battery. Increased visual blurring over baseline was reported by 5 subjects (13%) during the testing session. The median level of reported visual blurring did not increase significantly from baseline for the entire test battery. There was no difference between age groups in the median symptom increase from baseline.

Discussion

The aim of this study was to assess the test retest reliability, test stability, and concurrent validity of a gaze stabilization test in a healthy population of younger and older adults to determine whether the test is suitable for future study in patients with vestibular disorders. This study has shown that in a healthy population this version of the GST is reliable, causes minimal vestibular symptoms and measures a similar construct as the current clinical standard, the dynamic visual acuity test. Measurements of GST velocity and DVAT in this new version (i.e. with the InVision tunnel system) display good to excellent test retest reliability both within a single testing session and between sessions 7–10 days apart in subjects age 21–88 years. Measures of DVAT appear to be similarly reliable within and less reliable between sessions than measures of GST. Measurements of GST velocity were more stable (i.e. lower SEM%) across repeated testing both within and between sessions compared with DVAT in all subjects. Age did not appear to greatly influence test retest reliability within a single session for GST velocity. Inferences regarding age effect on GST reliability between sessions, however, are difficult in part due to the small sample size ($n = 10$) in each age group tested across 7–10 days. Estimates of test retest reliability for measurements of DVAT were good within a single session for both younger and older subjects. Gaze stabilization testing in the pitch plane was more stable for younger subjects as compared to older subjects whereas DVAT testing in yaw and pitch planes appeared slightly more stable in older subjects as compared to younger subjects.

A source of lower reliability in GST scores may be due to fatigue during GST assessments. Both the DVAT and GST use a fixed dwell time for the optotype to appear. The GST, however, is dependent upon the participant maintaining high head velocities throughout the trial, while the DVAT has a fixed velocity requirement of 60 or 85 degrees per second. The DVAT, therefore, may be associated with less fatigue than the GST.

The dynamic visual acuity test is currently used by physical therapists to assess the outcomes of rehabilitation in individuals with vestibular loss [15,16,24]. In the clinical dynamic visual acuity test, investigators refer to lines of visual acuity lost with head

movement rather than a raw dynamic visual acuity [19,24–26]. Though the version of the DVAT used in these studies differs in the presentation of the optotype from the version used in this study, the output of a loss of lines of visual acuity is the same. The output of the GST in degrees per second is an alternative measure for assessing the effectiveness of vestibular rehabilitation. Physiologic studies in humans have identified limits for eye head movement requirements during routine tasks such as ambulation [12,13,21]. Maximum VOR demands during walking for instance have been reported at 90 degrees per second and may be used as a clinical benchmark for individuals with vestibular dysfunction. While the GST scores in this healthy population far exceeded this threshold, future studies may identify VOR demands during more challenging tasks.

When comparing DVAT to GST, this study has found a moderate association between dynamic visual loss and GST performance. An explanation for why this relationship was not stronger is that the tests may be measuring similar, but different constructs. As mentioned briefly above, the number of failures reported during GST assessments indicates that participants are scoring incorrect responses for reasons other than deficiencies in the VOR. As a result, in healthy subjects who have no difficulty moving their head at high frequencies and large amplitudes, incorrect responses on the GST represent a poor VOR. In older adults whose head movements may be limited by pain or fatigue, failures are partly attributable to an inability to move the head rather than to VOR function.

There were particularly more trial failures in the older age group and in the pitch plane, possibly a result of increased prevalence of unmeasured symptoms such as neck pain [4] or fatigue due to muscle weakness [10] in the elderly population. These failures may in part account for the weaker correlation between GST and DVAT scores in the pitch plane, particularly for older adults. Additionally, since these trial failures count as incorrect responses, participants may be scoring worse because of these other symptoms. More failures in the pitch plane may therefore account for the lower GST scores seen in the pitch plane as compared to the yaw, a discrepancy not seen in the DVAT results.

While this study was not powered to detect a between group difference in GST scores of less than 70 degrees per second, it seems there was a trend to poorer performance in the older age group. This age effect was also seen for dynamic visual acuity. Prior studies have reported age-related decline in dynamic visual acuity [5,18]. Similarly, numerous studies have identified age-related decline in VOR performance [2, 3,8,9,20]. Pritcher et al. studied a prior version of the GST in a population of older adults and patients and found no group differences between patients with unilateral vestibular disease and older adults [22].

In this healthy population of younger and older adults, symptoms of nausea, visual blurring, and headache as reported by visual analog scale assessments were minimal after performing the GST and DVAT. There were minor symptoms of dizziness in some subjects and this should be recognized by technicians. A population of patients with vestibular disease who become dizzy with head movements more easily may experience more vestibular symptoms while undergoing testing with the GST or DVAT. Additionally, symptoms not measured in this study such as neck pain or neck muscle fatigue should be assessed in future studies.

An important limitation of this project is the selectivity of the sample. While the younger group may be more reasonably representative of healthy young adults, the older population may be healthier than a population based sample. There are numerous exclusion criteria including comorbid conditions and functional impairments that limit its applicability to older adults who commonly present to clinic visits with complaints of dizziness. While an effectiveness study of the ability of the GST to detect vestibular impairments in community dwelling older adults would be a useful undertaking, the intent of this project was to assess the reliability and concurrent validity of this measure in healthy controls before applying it generally, knowing that the reported values from this population will likely differ when applied to a general population of older adults.

Reliability of the computerized GST and DVAT is dependent on several parameters that may be difficult to control using parameter estimation (PEST) convergence algorithms. A potential improvement to the algorithm for the series of GST trials would be to present the same threshold for velocity on sequential trials to determine if an incorrect response was simply a trial failure or a true inability to identify the optotype at a required head velocity. An alternative approach to delivering the test battery would be to have a technician based exam similar to how audiologists administer audiograms in which the technician determines the level of testing based on subject response to the task. Trial failures due to inability to move one's head would be identified and labeled by a technician. This approach would allow the technician to evaluate whether the subject could not achieve a performance level on the GST because of an inability to move the head quickly or because they could not accurately identify the optotype while moving their head at the required velocity. While such an approach would require technical expertise, it would likely improve test reliability for wider use in a clinical setting.

Conclusion

The InVision system provides a compact way of standardizing prior versions of the GST and DVAT for outpatient clinical purposes. The GST causes minimal symptoms in a healthy population and has good test retest reliability, stability, and concurrent validity but could be improved with altered testing algorithms.

Acknowledgments

This project was supported by NIH grants DC005205, AG021885, AG024827, and AG10009. The authors would like to thank Anita Lieb and Susan Strelinski for technical assistance with this project.

References

1. Badaracco C, Labini FS, Meli A, et al. Vestibular rehabilitation outcomes in chronic vertiginous patients through computerized dynamic visual acuity and Gaze stabilization test. *Otol Neurotol*. 2007; 28:809–813. [PubMed: 17577127]
2. Baloh RW, Enrietto J, Jacobson KM, et al. Age-related changes in vestibular function: a longitudinal study. *Ann N Y Acad Sci*. 2001; 942:210–219. [PubMed: 11710463]
3. Baloh RW, Ying SH, Jacobson KM. A longitudinal study of gait and balance dysfunction in normal older people. *Arch Neurol*. 2003; 60:835–839. [PubMed: 12810488]
4. Binder AI. Cervical spondylosis and neck pain. *Bmj*. 2007; 334:527–531. [PubMed: 17347239]

5. Burg A. Visual acuity as measured by dynamic and static tests: a comparative evaluation. *J Appl Psychol.* 1966; 50:460–466. [PubMed: 5978038]
6. Demer JL, Honrubia V, Baloh RW. Dynamic visual acuity: a test for oscillopsia and vestibuloocular reflex function. *Am J Otol.* 1994; 15:340–347. [PubMed: 8579138]
7. Fleiss, JL. *Design and Analysis of Clinical Experiments.* New York: John Wiley & Sons; 1986.
8. Furman JM, Muller ML, Redfern MS, et al. Visualvestibular stimulation interferes with information processing in young and older humans. *Exp Brain Res.* 2003; 152:383–392. [PubMed: 12920495]
9. Furman JM, Redfern MS. Visualvestibular interaction during OVAR in the elderly. *J Vestib Res.* 2001; 11:365–370. [PubMed: 12446962]
10. Garces GL, Medina D, Milutinovic L, et al. Normative database of isometric cervical strength in a healthy population. *Med Sci Sports Exerc.* 2002; 34:464–470. [PubMed: 11880811]
11. Goebel JA, Tungsiripat N, Sinks B, et al. Gaze stabilization test: a new clinical test of unilateral vestibular dysfunction. *Otol Neurotol.* 2007; 28:68–73. [PubMed: 17106431]
12. Grossman GE, Leigh RJ, Abel LA, et al. Frequency and velocity of rotational head perturbations during locomotion. *Exp Brain Res.* 1988; 70:470–476. [PubMed: 3384048]
13. Grossman GE, Leigh RJ, Bruce EN, et al. Performance of the human vestibuloocular reflex during locomotion. *J Neurophysiol.* 1989; 62:264–272. [PubMed: 2754477]
14. Haley SM, FrigalaPinkham MA. Interpreting change scores of tests and measures used in physical therapy. *Phys Ther.* 2006; 86:735–743. [PubMed: 16649896]
15. Herdman SJ, Schubert MC, Das VE, et al. Recovery of dynamic visual acuity in unilateral vestibular hypofunction. *Arch Otolaryngol Head Neck Surg.* 2003; 129:819–824. [PubMed: 12925338]
16. Herdman SJ, Tusa RJ, Blatt P, et al. Computerized dynamic visual acuity test in the assessment of vestibular deficits. *Am J Otol.* 1998; 19:790–796. [PubMed: 9831156]
17. Leigh, RJ.; Zee, DS. *The Neurology of Eye Movements.* New York: Oxford University Press; 1999.
18. Long GM, Crambert RF. The nature and basis of age-related changes in dynamic visual acuity. *Psychol Aging.* 1990; 5:138–143. [PubMed: 2317293]
19. Longridge NS, Mallinson AI. The dynamic illegible Etest, A technique for assessing the vestibuloocular reflex. *Acta Otolaryngol.* 1987; 103:273–279.
20. Peterka RJ, Black FO, Schoenhoff MB. Age-related changes in human vestibuloocular reflexes: sinusoidal rotation and caloric tests. *J Vestib Res.* 1990; 1:49–59. [PubMed: 1670137]
21. Pozzo T, Berthoz A, Lefort L. Head stabilization during various locomotor tasks in humans. I. Normal subjects. *Exp Brain Res.* 1990; 82:97–106. [PubMed: 2257917]
22. Pritcher MR, Whitney SL, Marchetti GF, et al. The influence of age and vestibular disorders on gaze stabilization: a pilot study. *Otol Neurotol.* 2008; 29:982–988. [PubMed: 18667938]
23. Randolph, C. *Repeatable battery for the assessment of neuropsychological status (RBANS).* San Antonio, TX: Psychological Corporation; 1998.
24. Schubert MC, Herdman SJ, Tusa RJ. Vertical dynamic visual acuity in normal subjects and patients with vestibular hypofunction. *Otol Neurotol.* 2002; 23:372–377. [PubMed: 11981398]
25. Schubert MC, Migliaccio AA, Clendaniel RA, et al. Mechanism of dynamic visual acuity recovery with vestibular rehabilitation. *Arch Phys Med Rehabil.* 2008; 89:500–507. [PubMed: 18295629]
26. Schubert MC, Migliaccio AA, Della Santina CC. Dynamic visual acuity during passive head thrusts in canal planes. *J Assoc Res Otolaryngol.* 2006; 7:329–338. [PubMed: 16810569]
27. Taylor MM, Creelman CD. PEST: Efficient estimates on probability functions. *J Acoust Soc Am.* 1967; 41:782–787.
28. Tian JR, Shubayev I, Demer JL. Dynamic visual acuity during transient and sinusoidal yaw rotation in normal and unilaterally vestibulopathic humans. *Exp Brain Res.* 2001; 137:12–25. [PubMed: 11310166]
29. Whitney, SL.; Herdman, SJ. *Physical therapy assessment of vestibular hypofunction.* Philadelphia: FA Davis Company; 1999. p. 333-363.
30. Whitney SL, Marchetti GF, Pritcher M, et al. Gaze stabilization and gait performance in vestibular dysfunction. *Gait Posture.* 2009; 29:194–198. [PubMed: 18815040]

Fig. 1.

Scatter diagrams of Gaze Stabilization Test (GST) scores versus Dynamic Visual Acuity Test (DVAT) scores measured in (A) the pitch and (B) the yaw plane with 40 healthy subjects identified by age group stratified at 60 years: Results shown are average of two tests taken in one day.

Table 1

Mean (SD) Gaze Stabilization Test (GST) and Dynamic Visual Acuity Test (DVAT) results for three test sessions: Test 1 and test 2 on same day and test 3 7–10 days later

Group	Test	Test 1 mean(SD)	Test 2 mean(SD)	Test 3* mean(SD)
All subjects (<i>n</i> = 40)	GST (deg/sec)	Pitch 124 (32)	124 (35)	127 (32)
		Yaw 136 (39)	144 (40)	148 (32)
	DVAT (logMAR)	Pitch 0.16 (0.09)	0.14 (0.10)	0.15 (0.10)
		Yaw 0.17 (0.09)	0.16 (0.09)	0.15 (0.11)
Age < 60 years (<i>n</i> = 20)	GST (deg/sec)	Pitch 141 (25)	141 (30)	143 (25)
		Yaw 155 (36)	159 (38)	169 (21)
	DVAT (logMAR)	Pitch 0.14 (0.09)	0.15 (0.10)	0.18 (0.09)
		Yaw 0.16 (0.09)	0.13 (0.09)	0.15 (0.09)
Age ≥ 60 years (<i>n</i> = 20)	GST (deg/sec)	Pitch 108 (30)	108 (32)	112 (32)
		Yaw 118 (33)	129 (36)	123 (27)
	DVAT (logMAR)	Pitch 0.17 (0.09)	0.14 (0.11)	0.11 (0.11)
		Yaw 0.19 (0.09)	0.18 (0.09)	0.15 (0.13)

* *n* = 20 (10 in each age group) received test 3.

Table 2

Reliability estimates (Intraclass correlation coefficients with 95% confidence intervals) for pitch and yaw plane GST and DVAT measurements taken within and between testing session for all subjects and age groups (age < 60 years, age $\geq$ 60 years)

Measurement/Plane	All subject		Age < 60 years		Age $\geq$ 60 years	
	Within session (n=40)	Between session (n = 20)	Within session (n = 20)	Between session (n = 10)	Within session (n = 20)	Between session (n = 10)
GST/Pitch	0.69 (0.48–0.82)	0.54 (0.14–0.79)	0.66 (0.31–0.85)	0.52 [*]	0.52 (0.11–0.78)	0.34 [*]
GST/Yaw	0.75 (0.58–0.86)	0.59 (0.21–0.81)	0.62 (0.25–0.83)	0.22 [*]	0.79 (0.54–0.91)	0.57 [*]
DVAT/Pitch	0.60 (0.36–0.77)	0.10 [*]	0.57 (0.19–0.81)	0.26 [*]	0.64 (0.29–0.84)	0 [*]
DVAT/Yaw	0.56 (0.30–0.74)	0.49 (0.08–0.76)	0.51 (0.10–0.77)	0.52 [*]	0.56 (0.17–0.81)	0.47 [*]

* $p > 0.05$, Analysis of Variance F-test against ICC = 0.

Table 3

Standard error of measurement (SEM) as estimates of within subject variability and minimal detectable change (MDC) for measures of GST/DVAT within and between testing sessions for (A) all subjects and for age groups (B) age < 60 years and (C) age ≥ 60 years

A.									
Measurement/plane	All subjects								
	Within session				Between session				
	SEM*	SEM% [‡]	MDC ₉₅ [‡]	SEM*	SEM% [‡]	MDC ₉₅ [‡]	SEM*	SEM% [‡]	MDC ₉₅ [‡]
GST/Pitch	18.54	14.9	51.39	19.60	15.0	54.33			
GST/Yaw	19.70	14.1	54.61	19.21	13.0	53.25			
DVAT/Pitch	0.06	35.6	0.16	0.09	53.4	0.24			
DVAT/Yaw	0.05	26.5	0.12	0.06	37.4	0.18			

B.									
Measurement/plane	Age < 60								
	Within session				Between session				
	SEM*	SEM% [‡]	MDC ₉₅ [‡]	SEM*	SEM% [‡]	MDC ₉₅ [‡]	SEM*	SEM% [‡]	MDC ₉₅ [‡]
GST/Pitch	15.86	11.3	43.96	15.80	10.9	43.79			
GST/Yaw	22.81	14.1	63.22	19.43	11.8	53.86			
DVAT/Pitch	0.06	42.2	0.16	0.08	55.3	0.21			
DVAT/Yaw	0.06	39.4	0.17	0.06	39.0	0.17			

C.									
Measurement/plane	Age ≥ 60								
	Within session				Between session				
	SEM*	SEM% [‡]	MDC ₉₅ [‡]	SEM*	SEM% [‡]	MDC ₉₅ [‡]	SEM*	SEM% [‡]	MDC ₉₅ [‡]
GST/Pitch	21.20	19.7	58.76	23.15	19.7	64.18			
GST/Yaw	15.76	12.8	43.70	18.30	13.9	50.71			
DVAT/Pitch	0.05	31.8	0.15	0.09	52.9	0.25			
DVAT/Yaw	0.06	31.4	0.17	0.07	34.5	0.18			

* SEM in original units of measurement for GST (deg/sec) and DVAT (logMAR).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

$\sqrt{\text{SEM}}\% = \text{SEM}/\text{Mean}$.
 $\sqrt{\text{MDC}}95 = \text{amount of change attributable to measurement error}$.