ARTICLE

Reliability and Validity of a Computerized Tachistoscope Test

John "JT" Tassinari, OD, FAAO, FCOVD
Western University of Health Sciences College of Optometry, Pomona, California; private practice Pasadena, California

ABSTRACT

This study evaluates the reliability and validity of the Perceptual Therapy Systems (PTS) tachistoscope test. 74 poor readers in grade school were tested twice with the tachistoscope test. This test purports to measure visual skils important in reading. Poor readers should perform poorly on it.

Results: The PTS tachistoscope has good-to-excellent reliability per the intra-class correlation coefficient (ICC) and fair-to-good reliability per Cohen's Kappa index. The poor readers performed, on average, at the scaled score level of 7.

Conclusions: The PTS tachistoscope test has sufficient reliability. Its validity is supported by the fact that poor readers, determined as such by case history, performed poorly on it.

Correspondence regarding this article should be emailed to John "JT" Tassinari, OD, at jtassinari@westernu.edu. *All statements are the authors' personal opinion and may not reflect the opinions of the College of Optometrists in Vision Development, Vision Development & Rehabilitation or any institution or organization to which the authors may be affiliated. Permission to use reprints of this article must be obtained from the editor.* *VDR is indexed in the Directory of Open Access Journals. Online access is available at www.covd.org. https://doi.org/10.31707/VDR2016.2.4.p244*

Tassinari JD. Reliability and validity of a computerized tachistoscope test. Vision Dev & Rehab 2016;2(4):244-50.

Keywords: tachistoscope, test-retest reliability, vision therapy evaluation,

INTRODUCTION

When the eyes are stationary, our vision can identify and process discrete parcels of information during very brief intervals. This capability serves the act of reading well. A reader can pause the eyes for a short moment, on the order of ¼ second, and glean useful information from the text before transiting to a new parcel of text with a saccade. A fixation pause while reading can be described in terms of its content (e.g. 12 characters of text); spatial extent of the content (eg 5 degrees of arc) and actual time spent paused. More content in a wider span, accurate visual processing of content, and speedier pauses make reading more efficient and vice versa. The content of a fixation pause can be delineated into direct view, clear text that is attended to in central vision, and peripheral preview which contains blurred parafoveal text in the right hemifield. Figure 1 shows this delineation along with a depiction of the typical point of fixation.

The visual processing that takes place during a fixation pause includes pattern recognition of fixation pause contents in direct view and pattern recognition plus spatial processing of peripheral preview. The pattern recognition/spatial processing in the peripheral preview primes the visually guided motor plan for the next saccade. Good spatial processing locates the optimal landing site for the next fixation pause which is generally located between the beginning and middle of the target word.[1] The distance from one optimal landing site to the next varies because word lengths vary. Thus, efficient readers will vary their saccade length based on visual information in the peripheral preview. If the optimal landing site is consistently hit by an accurately programmed and executed saccade, the fixation pause is quicker and there are fewer re-fixations.[2,3] The pattern recognition portion of the peripheral preview serves another purpose. It provides a blurry peek at the text expected to be in next fixation pause direct view. This peek may reveal a small word such as "to" or spaces and the reader

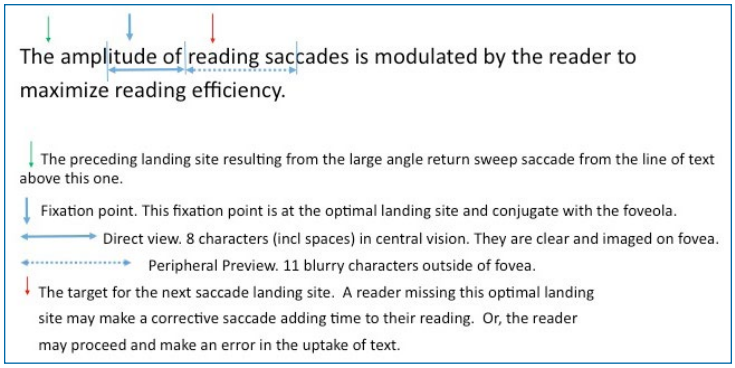


Figure 1. Contents of a fixation pause while reading are direct view + peripheral preview

may program a longer saccade to skip them.[2] The peek may also lead to partial or complete visual recognition of a high frequency word (eg "this") and the ensuing fixation pause will be quicker.[2,3] Another visual task associated with a fixation pause is continued visual processing of the fixation pause content after its conclusion.[4] This continued visual processing is important because the printed symbols of language in reading may require more time to process than is available in a fixation pause.[2,5] Additionally, a complete word may not fit in the fixation pause direct view. For example, in Figure 1 the last 5 letters in the word 'amplitude' extend beyond the border of the previous fixation pause direct view. The image 'a-m-p-l' is retained across the ensuing saccade to match it up to the next fixation pause which, if the saccade is accurate, will contain 'i-t-u-d-e'. To sum up, a multitude of simultaneous or nearly simultaneous visual functions are called upon during a fixation pause to aid the task at hand, decode the text, and plan an accurate saccade to the next segment of text. A term for the visual perceptual skills that underlie the pattern recognition and spatial processing that take place during a fixation pause is speed and span of visual perception.

Optometrists who evaluate visual function above and beyond refractive status consider speed and span of visual perception to be a learned, developed skill that can be evaluated and, if necessary, this skill can be trained and improved with optometric vision therapy (VT). A common method of testing speed and span of visual perception is Tachistoscope.

A tachistoscope is an instrument which exposes visual stimuli for a defined brief period of time.[6] An individual viewing a tacshistoscopically presented target receives the information during the allotted time span and then sustains the intake long enough to plan and activate a response. If the intake is incomplete or if the sustain step post intake is tenuous, errors in the response are predestined. Renshaw conducted the pioneering work with tachistoscope in the early and mid-parts of the previous century.[7] His early work emphasized training vision with a tachistoscope for purposes of rapid recognition of combat aircraft and later reading efficiency.[7] Solan and colleagues further researched tachistoscope to establish test norms and show that there is a relationship between the visual skills measured by tachistoscope and academic performance.[8-13] This research is the basis for the inclusion of tachistoscope testing in an optometric work-up of a child with learning problems. Solan described the skills measured by tachistostopic testing in this way.[9]

"The test provides a measure of the child's visual memory, visual processing, and degree of perceptual accuracy and automaticity."

The research providing norms for the tachistoscope test utilized a tachistoscope ("Tach-X, Educational Development Laboratories) that projects the test targets (numbers) 8 feet away.[8-13] A computerized Tachistoscope Test (TT) was introduced by Groffman as part of a computer system called RCT[14] then later a software program called Perceptual Therapy Systems ("PTS") Test.[15] A computer has the advantages of reproducible speed, target size, and luminance. The purpose of the present study is to investigate the test-retest reliability of the Tachistoscope Test (TT) that is part of the PTS test battery. There have been no prior published studies of this kind. Reliability is important to support the veracity of diagnosis and treatment effect.[16] A test that is unreliable

may result in a different score after treatment because it fluctuated as opposed to a true change due to treatment effect. A lesser component of the study will be validity. The validity research question is whether or not the group tested (poor readers per case history) performs poorly on TT.

METHODS

Seventy-four subjects were enrolled in the study as they consecutively presented to the authors optometry office. All 74 had the following characteristics to qualify as subjects for this study.

1. The patient presented requesting and expecting a work-up to determine if VT is indicated.
2. Difficulty reading was an expressed concern in the case history.
3. There was no active eye disease requiring treatment or referral.
4. Binocular near acuity was 20/30 Snellen or better.

In the author's practice, testing conducted during a VT work-up is typically divided between two office visits. The 74 subjects completed all testing in the usual manner other than having TT administered on the first office visit and repeated on the second office visit. In all cases, the subjects wore the same lenses for each TT. If a new lens prescription that would normally be worn for nearpoint was determined at office visit 1, it was placed in a trial frame for TT. At office visit no. 2, the new prescription was in place in an ordinary dress frame. The same computer monitor and keyboard was used for all TT administrations.

TT was administered according to the test instructions provided in the software help section. The examiner enters the patient name and importantly for scoring purposes, the age in years. The PTS TT has a practice mode to orient the patient to the test and gain familiarity with the keyboard. In brief, the test exposes a string of numbers in an escalating sequence for 0.10 second. A 1- digit target is flashed to begin the test regardless of age. There are 2 trials for each sequence length. That is, 2 trials that are 1-digit, 2 trials that are 2-digits, 2 trials that are 3-digits etc. The patient views the flashed numbers and types in a response. After typing in the response, the patient may change her/his answer by backspace deleting a number. If no corrections are made, patient depresses the space bar once, and then depresses it again causing the computer to present the next number string. If the response to 2 consecutive same length test items are incorrect, the test ceases automatically and scores are provided by the software. The scores are raw, scaled, and percentile. The raw scores are a product of the number of correct answers before two consecutive errors. If presentations 1 – 5 were correct, for example, followed by an error on presentation 6, correct on presentation 7 then 2 consecutive errors, the raw score would be 6 points. The software also provides a rating of the score (eg "a little below average") that is based on a previously published rating scale.[14]

All subjects received a practice session during which the subject gained familiarity with the keyboard and where to look on the monitor. During the practice session each subject was given explicit verbal instructions and a demonstration that deleting a keyboard entry with the backspace key and then re-entering another number(s) was permissible. Following the practice session, study subjects interfaced with only the computer while the examiner was silent. This test taking mode reduced examiner bias. When the second TT was administered on visit 2, the examiner and subject were blinded to the results of the first TT.

DATA ANALYSIS

At the conclusion of the study, the raw score and scaled score for each subject from TT1 and TT2 were entered into a spread sheet for statistical analysis. Each subject was rated in terms of pass or fail TT. A scaled score of 8

or lower was considered a fail as is the author's practice and consistent with a published rating scale.[16,17] Measures of central tendency were calculated along with intra-class correlation coefficient and Cohen's Kappa Index.

RESULTS

The 74 subjects ranged in age from 7 to 18 years, had an average age of 10.9 years (sd 6.3), and there were 34 females. They were unevenly distributed among grades 1 – 12 with 3rd – 5th graders making up 50% of the study population (Table 1). The number of days between TT1 and TT2 averaged 18 with a range of 1-46. Fifty-eight of the subjects had a fail score (scaled score 8 or lower) at TT1. Table 2 shows the mean scaled score for the entire group was almost exactly 7 at TT1 and TT2. The mean raw score was likewise very close to unchanged at 6.3 on TT1 and 6.6 on TT2. Thus, this group of schoolchildren who had reading problems per parental report performed poorly on TT.

Table 1. Distribution of study population by grade

Grade	Number %
1	6 (8.1)
2	7 (9.5)
3	15 (20.2)
4	13 (17.6)
5	9 (12.2)
6	6 (8.1)
7	8 (10.8)
8	2 (2.7)
9-12	8 (10.8)
Total	74 (100%)

Test-retest reliability was assessed via the intra-class correlation coefficient (ICC) and Cohen's Kappa index. ICC is an overall index of reliability that may range from 0 (no reliability) to 1 (perfect reliability).[18] Cohen's Kappa index (K) measures to what extent test and retest agree with regards to pass fail.[19,20] A K of 0 means complete disagreement. K of 1 means zero change in pass-fail classification upon retest. The range of ICC and K values is commonly interpreted as follows:

ICC or K <. 4	0.4 < ICC or K < 0.75	ICC or K > 0.75
Poor reliability	Fair-to-good reliability	Good-to-excellent reliability

Table 2 Reliability indices plus mean raw scores and mean scaled scores from Test Administration 1 (T1) and T2

TT Administration	Raw Score mean (sd)	ICC,[a] Raw Score (95% CI)	Scaled Score mean (sd)	Kb, Scaled Score
T1	6.28 (1.92)	0.80 (0.71, 0.87)	6.99 (2.28)	0.48 (p <.001)
T2	6.61 (2.01)		7.32 (2.61)	

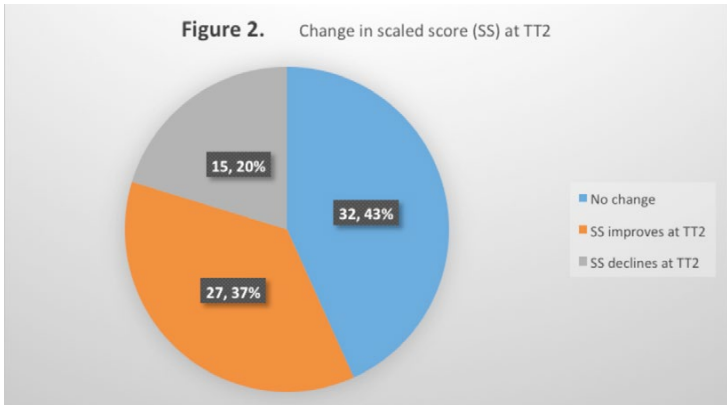
Table 3. Separation of subjects into borderline scores, scaled score (SS) = 9 or 8, and non-borderline scores (SS > 9 or SS < 8)

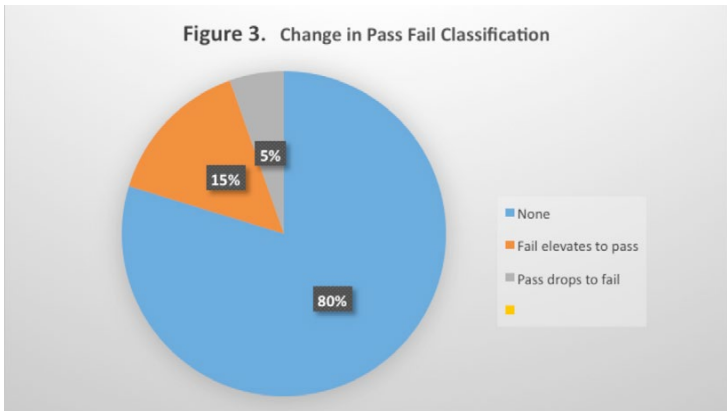
	Pass[a] TT1 n = 16	TT 2, Pass drops to fail n = 4		Fail TT1 n = 58	TT2, Fail elevates to Pass n = 11
SS = 9	10	4	SS = 8	22	9 (41%)
SS > 9	6[b]	0	SS < 8	36[b]	2 (5.5%)

a, Pass: scaled score (SS) ≥ 9
b, K for these 42 non-borderline subjects = 0.83 (p < 0.001)

Table 2 shows the ICC for the raw score was 0.804 (95% ci, 0.706, 0.872) placing it in the good-to-excellent category with the lower end of the 95% confidence interval very close to the good-to-excellent category. K for the scaled score (Table 2) was 0.484 for the entire study population placing it in the fair-to-good reliability category. A closer look at the study population in terms of pass-fail classification reliability shows that subjects with scaled scores on the cusp of the pass-fail cutoff (SS = 8 or 9) were less reliable and vice versa. As shown in Table 3, none of the subjects who passed with a scaled score of 10 or higher (n = 6) dropped to fail on the second administration. Among the 36 subjects who failed TT on the first administration with a scaled score of 7 or lower, 2 elevated to pass on the second administration. This group of 42 non-borderline subjects had a very good Cohen's Kappa index of reliability of 0.83 (Table 3) placing it well into the good-to-excellent descriptive category.

Some qualitative information about the change in scores TT1 versus TT2 is shown in Figure 2 and Figure 3. Figure 2 shows that zero change in scaled score was the most common outcome followed closely by 37% of the





subjects demonstrating an increase in scaled score upon retest. Least common was a drop in scaled score upon retest.

DISCUSSION

The results of this study can be analyzed in terms of its reliability, with particular attention to error analysis and case analysis, and in terms of its validity.

Reliability

The results of this study show that the tachistoscope test in the Perceptual Therapy Systems software has good-to-excellent reliability per the ICC statistical analysis. It has fair-to-good reliability per the Cohen's Kappa index statistic. It is extremely reliable for patients with a clear cut fail (SS < 7) or clear cut pass (SS > 10) as evidenced by a Kappa score of 0.83 (p-value < 0.001) for this group. Optometrists can include it in a VT workup with confidence. Two features of TT contribute to its reliability. First, a generous practice session precedes the actual test to move the patient past the learning curve. The practice session may also minimize keyboard errors during the test. Second, the patient interfaces with the computer, not the examiner, once the test begins and therefore fluctuations in the way the examiner gives the test do not come into play. Third, the computer based format can be counted on to present the test items the same way each time.

TT does not have perfect reliability as expected because it is a behavioral performance test. It is particularly vulnerable to a change in pass-fail rating when the score is on the cusp of the pass-fail cutoff. This vulnerability makes sense when one considers that a small fluctuation in raw score near the cusp can cross the pass-fail border. A similar fluctuation of a raw score well below or well above the pass-fail cutoff will not cross it. Thus, borderline scores place a premium on error analysis and case analysis. A brief discussion of error analysis and case analysis follows.

Error Analysis & Case Analysis

The 2 types of errors, false low and false high, are shown in Figure 3. Errors 'A' and 'B' do not require analysis because they would not occur in the typical clinical scenario of administering TT once. They are relevant to a study such as this one. However, there is no way to know which type of error occurred.

	First score (T1) is low and elevates to normal or high upon retest (T2)	Normal or high score drops to low upon retest
First Score is True	A T2 is false high	B T2 is false low
Second Score is True	C T1 is false low	D T1 is false high

Figure 4. Four possible errors if test is administered twice. Assumes no treatment or change in the patient in the interval between tests.

Error 'C' is the one clinicians probably deliberate and encounter most often. It would occur during administration of TT (and other performance tests) if the patient did not put forth good effort and attention or if patient did not understand the instructions. In this type of error, the skill is better than depicted by the test score. Adroit and attentive clinicians approach "C" errors in 2 ways. First, they give instructions and adopt

a demeanor during testing that engenders good effort and attention on the part of the patient toward a goal of preventing false low errors. Second, they utilize their powers of observation[21] to conclude a false low score occurred. Upon doing so, they may repeat the test or apply high level case analysis (see below). Error 'D', false high, is a clinical conundrum, impossible to know if it occurred, and, fortunately, uncommon. The skill is truly low but the test dubiously represents it as normal and VT may be incorrectly withheld. A false high error may occur by luck or chance. Tests with a finite set of answers are prone to this type of error if the patient guesses correctly. TT is such a test. It purportedly measures speed and span of perception but the actual behavior measured is patient depression of keys on a keyboard. Patients taking TT have a 10% chance of selecting the correct number by guessing. By comparison, a performance test which is a direct measure of the skill is not prone to chance. For example, the Developmental Test of Visual Motor Integration is unlikely to result in a false high score because the patient luckily drew one of the test patterns correctly. One more example to illustrate the likelihood of a false low error versus a false high error can be found in visual field testing. Abnormal visual field test results are common among individuals with truly normal visual fields. However, it is difficult to imagine a patient with a true hemianopsia luckily pushing the response button just after an unseen test point to pass the test and hide their problem causing a false pass error.

One strategy for dealing with the uncertainty surrounding a borderline test score is to repeat the test. This strategy may be impractical because it could lead to a time-consuming repeat of manifold tests in a VT work-up. A potentially powerful case analysis method for deciding if a low score is true or false is to cull the case history for the symptoms and performance concerns likely to be present if the skill is truly low.[17,21,22] A low score coinciding with expected symptoms adds credence to the diagnosis of visual dysfunction and vice versa. Slow and error-prone word recognition for common words are symptoms that may be present if TT is used to diagnose deficient speed and span of visual perception.[8,14,21] Another strategy is to view the score in the context of the entire workup. If other like tests match the results of TT, the TT score can be viewed as a correct representation of the patients speed and span of visual perception. For example, a TT scaled score of 8 (25th) percentile accompanied by similar scores on visual memory tests such as Test of Visual Perceptual Skills Visual Memory subtest adds credence to the TT score. This method of grouping like tests, cluster analysis, can be carried out with more than one filter. Another filter that is appropriate for TT is the simultaneous versus successive supramodal processing dichotomy espoused by Kirby/Das[23] and Groffman/Solan.[14] Visual perceptual tests that require simultaneous/global "all-at-once" coding and integration of information are considered simultaneous.[14] Spatial characteristics of the visual information are important. In simultaneous processing, the location and features of the visual information is surveyable all at once.[14] TT is therefore a test of global, simultaneous processing. If the patient shows a pattern of scoring low on visual perceptual tests that fall into the simultaneous category, a borderline low score on TT can be viewed as truly low. The opposite is also true. If the patient is generally good with simultaneous processing then doubt is cast on the borderline score and it may be reasonable to conclude a false low is present. Another cluster within which TT results may be analyzed is visual processing speed.[14] To sum up, the results of TT will influence the clinical algorithm used to conclude visual dysfunction is present, or not, and, whether or not VT will be prescribed. A score that is a clear cut fail or pass, matches the symptom status, and fits a cluster will have stronger influence.

Validity

If TT measures a visual perceptual skill that is part and parcel of reading efficiency then a group of poor readers should score low on it.

This population of poor readers did exactly that. The average scaled score for the group was approximately 7 (16th percentile) lending validity to TT. Limitations to this conclusion are that poor reading was determined by self-report on case history rather than by standardized reading assessment. Another limitation is that there was no comparison group of good readers. It is not known from this study if good readers would perform better than poor readers on TT.

CONCLUSIONS

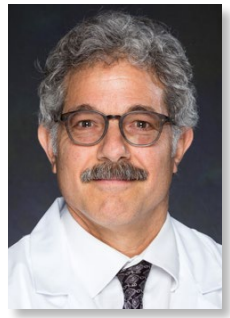
The computerized tachistoscope test of speed and span of visual perception in the PTS test battery is reliable, especially for scores that are clear-cut pass or fail. Scores that are near the pass-fail cutoff should be viewed with some caution. Suggestions for case analysis of borderline PTS tachistoscope test scores are provided that may be generalized to other performance tests in a vision therapy workup. The PTS tachistoscope test is valid insofar that a group of poor readers, determined per case history, performed poorly on 2 administrations of it. Limitations to this conclusion are that poor reading was determined by self-report on case history rather than by standardized reading assessment.

Conflict of Interest

The author has no financial interest or affiliation with Home Therapy Systems and the sale of Perceptual Therapy Systems Tachistoscope test.

REFERENCES

1. Rayner, K., Fischer, M. H., & Pollatsek, A. Unspaced text interferes with both word identification and eye movement control. Vision Research 1998; 38: 1129–1144.
2. Rayner K. Eye movements in reading and information processing: 20 years of research. Psychol Bull 1998; 124 (3): 372-422.
3. Lavigne F,Vitu F, d'Ydewalle G. The influence of semantic context on initial landing sites in words. Acta Pcychol 2000; 104: 191-214.
4. Boden C, Brodeur DA. Visual processing of verbal and nonverbal stimuli in adolescents with reading disabilities. J Learn Disabil 1999; 32 (1): 58-71.
5. Gilbert LC. Speed of processing visual stimuli and its relation to reading. J Educ Psych 1959; 55 (1): 8-14.
6. Dictionary of Optometry and Visual Science, 6th ed. Millodot M. Boston: Butterworth-Heinemann, 2004.
7. Renshaw S. The visual perception and reproduction of forms by tachistoscopic methods. The J of Psych 1945; 20: 217-232.
8. Solan HA. Visual processing training with the tachistoscope: a rationale and grade one norms. J L Disabil 1969; 2(1):32-8.
9. Solan HA, Mozlin R, Rumpf DA. Selected perceptual norms and their relationship to reading in kindergarten and the primary grades. J Amer Optom Assoc 1985;56 (6):458-66
10. Solan HA, Mozlin R. The correlations of perceptual-motor maturation to readiness and reading in kindergarten and the primary grades. J Amer Optom Assoc 1986; 57 (1):28-35.
11. Solan HA. The effects of visual-spatial and verbal skills on written and mental arithmetic. J Amer Optom Assoc 1987; 58 (2): 88-94
12. Solan HA. Perceptual norms in grades 4 and 5: a preliminary report. J Amer Optom Assoc 1987; 58 (12): 979-82.
13. Krumholtz I, Solan HA. Selected perceptual norms for the learning-disabled population ages 6 through 9. J Beh Optom 1993; 4 (2): 31-36.
14. Groffman S. Solan HA. Developmental and perceptual assessment of learning-disabled children. Santa Ana: Optometric Extension Program Foundation, 1994.
15. Home Therapy Solutions, Inc. http://bit.ly/2hUGvYe
16. Solan HA, Suchoff IB. Tests and measurements for behavioral optometrists. Santa Ana: Optometric Extension Program Foundation, 1991
17. Tassinari JT. Assessing the assessment: learning related vision problems test scores revisited. Optom Vis Develop 2008; 39 (3): 128-39.
18. Fleiss JL. The design and analysis of clinical experiments. New York: Wiley, 1986.
19. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33: 159-174.
20. Newman TB et. al. Designing studies of medical tests. In Hulley SB et al (eds) Designing clinical research 3rd ed. Philadelphia: Lippincott, Williams, and Wilkins, 2007: 183-205.
21. Hoffman LG, Rouse MW. Referral recommendations for binocular function and/or developmental perceptual deficiencies. J Amer Optom Assoc 1980; 51 (2): 119-125.
22. Borsting E. Visual perception and reading. In Garzia RP (ed). Vision and reading. St. Louis: Mosby, 1995: 153-54.
23. Kirby JR, Das JP. Reading achievement, IQ, and simultaneous-successive processing. J Educ Psychol 1975; 69: 564-70.



AUTHOR BIOGRAPHY:
John "JT" Tassinari, OD, BS, FAAO, FCOVD

- OD 1987, Southern California College of Optometry
- Residency 1988, SUNY College of Optometry
- Private Practice, Pasadena, California
- FCOVD 1993
- Diplomate AAO 2006
- Chief of Pediatric Optometric Service and Associate Professor, Western U. of Health Sciences