Assessing the Assessment: Learning Related Vision Problems Test Scores Revisited

JT Tassinari, OD FCOVD
Private Practice, Pasadena, CA

The results from the Visual Efficiency (VE) and Visual Information Processing (VIP) tests that comprise an Optometric Vision Therapy (VT) work-up of an individual with learning problems require careful analysis. An important step in the analysis is rating the test results. For nearly all VIP tests and many VE tests, the raw test score (e.g. points, time elapsed, cycles per minute, number of errors) is converted to a performance score such as percentile rank. This performance score is rated as pass or fail or along a continuum of descriptors such as poor, below average, average, and above average. The obvious implication is that low performance scores equal visual dysfunction and that VT should be prescribed. Optometrists administering VE and VIP tests are faced with the important decision of selecting a scale that rates the performance score. So, how low does a performance score have to be to be considered a fail?

The Optometric Extension Program (OEP) formulated a table of expected values for phoropter based tests of visual function that was first published in 1935. According to Birnbaum, the expecteds “…are values assumed to be the minimum required if an individual is to withstand the impact of nearpoint stress and maintain satisfactory performance.” Birnbaum further states that the OEP expected values are not normative (i.e. based on population statistics) but, “…are findings that are theorized to be optimal for a given population.” In the OEP case analysis system, findings are compared to the expected value and rated as high or low. The OEP expected values later proved to be very close to mean values determined from population distribution studies. In practice, the OEP analytical system rates test results as low if they are just below the mean.

Nine years later, in 1944, Morgan published the mean and standard deviation for phoropter based accommodation and vergence tests. He recommended that the range of normal be set at 0.50 standard deviations (SD) below and above the mean. Thus 0.6 SD below the mean would be a low score. In Morgan’s system, a raw test score has to be further below the mean to be considered low than in the OEP system. For example, the OEP expected for the PFC break at near (#16B) is 21BO. Morgan reports the mean PFC near break to be 21BO with a SD of +/- 6 prism diopeters. In the OEP system, 20BO is a low finding. In Morgan’s system, 18 BO is in the normal range so the finding is not low until it reaches 17BO or lower. Griffin and Grisham’s text uses Morgan’s data and 0.6 SD below the mean criterion. They provide tables that convert the test result directly to a description. For example a near PFC break of 20 is converted to “adequate.” Seventeen is converted to “weak.”

Fifty four years after Morgan, the Optometric Clinical Practice Guideline (CPG) ‘Care of the Patient with Accommodation and Vergence Dysfunction’ recommends “…any finding that deviates from the norm by 2 standard deviations may indicate an anomaly.” Per this criterion near PFC Break has to be 9BO before it is considered anomalous if Morgan’s mean of 21 and SD of 6 is used. Elsewhere in this CPG and contrary to the 2 SD criterion, a table of normative values for phoropter based accommodative and vergence tests is provided that uses 0.6 or 0.7 SD below the mean as low.
Regarding VIP test results, an authoritative and up-to-date answer to the question of how low a score has to be to be considered a fail can be found in the Optometric Clinical Practice Guideline (CPG) titled ‘Care of the Patient with Learning Related Vision Problems.’ This American Optometric Association document provides the following recommendation for VIP test results.

A test result with a z score that is ≥ 1.5 standard deviations below the mean (percentile rank = 6.68) should definitely be considered anomalous and clinically significant. Scores falling between 1.0 and 1.5 standard deviations below the mean should be considered suspicious and perhaps clinically relevant, depending on the overall clinical picture, the nature and type of the learning problem, and the level of overall cognitive function.

As I write this article, there are five learning related vision problem (LRVP) patients enrolled in a VT program at my office whose entire set of VIP performance scores are above the 7th percentile. Two of these have all VIP test scores above the 16th percentile. There are seven other patients enrolled in VT who have individual VIP skills that were rated as deficient yet all of the tests administered for that skill resulted in scores that were above the anomalous criterion provided in the LRVP-CPG. Importantly, all of these patients have symptoms that are associated with the VIP skill(s) diagnosed as deficient. According to the LRVP-CPG scale, VIP VT was prescribed to as many as twelve children with normal VIP function.

I have concerns with the LRVP-CPG Scale because the anomalous criterion for VIP test results is too low. The purpose of this article is to discuss some of the statistical and theoretical underpinnings of setting a rating scale for the performance test scores from a VT work-up. I also present information on symptomatology and case analysis as they relate to a rating scale and review rating scales used outside of optometry. Finally, I suggest an alternative to the LRVP-CPG scale that is based on previously published optometric rating scales and integrated with case analysis.

Statistical Considerations

There are three statistics that are elemental to establishing a rating scale. The first two are mean and standard deviation (SD). Mean is the arithmetic average of all the raw test scores in a population. A liberal rating scale would use the mean as the expected result and any score below the mean would be considered anomalous. SD is a mathematical expression of how the raw test scores are distributed about the mean in a normal bell-shaped distribution. A high SD means there are raw scores spread widely above and below the mean. The basis for all performance scores (derived scores) is discrepancy between raw test score and mean in SD units. A “Z” score, the third relevant rating scale statistic, is a way to convert a raw score to its difference from the mean in SD units. The formula is

\[ Z = \frac{\text{Raw Score} - \text{Mean}}{\text{SD}} \]

If the raw test score equals the mean then Z=0. If the raw score is exactly one SD below the mean then Z = -1.00.

Once Z is calculated, the examiner can convert Z to a performance score term of her/his choosing. The trend in optometry is to use percentile rank and Table 1 shows percentile rank and three other performance score terms.

The main benefit of a Z score (and converting it to a performance score such as percentile) is that it makes a raw test score meaningful. Consider a 4th grader whose Developmental Eye Movement (DEM) Ratio Score is 0.10 higher (worse) than the mean and a 7th grader who also scores 0.10 above the mean (high DEM ratio score means worse performance.) The 7th grader has a higher Z score and thus a lower performance score (10th percentile) than the 4th grader (23rd percentile). Why? The SD for 7th graders is smaller than 4th graders. In other words, 0.10 Ratio points is further from the mean in SD units for 7th graders than 4th graders.

One final comment regarding statistics is that optometrists, particularly those prescribing VT, have a tradition of converting the actual test score

<table>
<thead>
<tr>
<th>Z Score</th>
<th>Percentile Rank</th>
<th>Scaled Score</th>
<th>Standard Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.5</td>
<td>7th</td>
<td>5.5</td>
<td>78</td>
</tr>
<tr>
<td>-1.0*</td>
<td>16th</td>
<td>7</td>
<td>85</td>
</tr>
<tr>
<td>-0.5</td>
<td>31st</td>
<td>8.5</td>
<td>93</td>
</tr>
<tr>
<td>0.0*</td>
<td>50th</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>+0.5</td>
<td>69th</td>
<td>11.5</td>
<td>107</td>
</tr>
<tr>
<td>+1.0*</td>
<td>84th</td>
<td>13</td>
<td>115</td>
</tr>
<tr>
<td>+1.5</td>
<td>93rd</td>
<td>14.5</td>
<td>122</td>
</tr>
</tbody>
</table>

a, raw score is 1 SD below mean  
b, raw score = mean  
c, raw score is 1 SD above mean
(PRA/#20 = -1.25) directly to a rating of the score (low) as opposed to converting to Z or percentile rank first, then rating the performance score. This tradition should not preclude the optometrist from understanding and applying basic statistical concepts to their visual skill testing. For a comprehensive yet concise review of the statistical concepts involved in visual skill testing the reader is referred to *Tests and Measurements for Behavioral Optometrists.*

**Theoretical Considerations**

The optimal scale for a test of visual function would designate a visual skill as problematic if, and only if, it is causing symptoms/performance problems for that patient. This optimal scale is based on whether or not the consequences of visual skill dysfunction are present, not population distribution statistics. It can be conceptualized in the epidemiological terms of specificity and sensitivity. Although specificity and sensitivity are usually applied to screening tests, Hoppe discusses their utilization in diagnostic tests.

Figure 1 pits the visual skill test result against the true state of the patient’s visual function. A test and rating scale with perfect specificity would always result in pass for normal patients. There would be no False/Negative (‘F/N’) results as shown in box 2 of Figure 1. Perfect sensitivity means the test is 100% sensitive to visual dysfunction. The test score always results in a rating of fail if the patient has symptomatic visual dysfunction. The P/VD error shown in box 3 of Figure 1, passing a patient who should have failed, would not occur.

Figure 2 shows a stylized distribution of visual function along a percentile rank scale. It shows that specificity and sensitivity are influenced in opposite directions by the placement of the anomalous/normal criterion. If the placement is too low, as is the LRVP-CPG criterion in my opinion, specificity suffers while sensitivity profits. This low placement prevents the problem of prescribing VT to individuals who do not need it. However, it causes a problem of VT denied to individuals with visual dysfunction. A criterion placed too close to the mean, C₂ in Figure 2, results in the exact opposite.

In discussing their research on normative data for the Rosner Test of Visual Analysis, Rateau et al. make this same point.

Setting the threshold too high will result in too many children receiving remediation unnecessarily but few children missing out on remediation who are likely to require it; setting threshold too low will result in some children not receiving remediation when they most likely require it, but few children receiving it unnecessarily.

Figure 2 also depicts a range of test scores, between C₁ and C₂, in which there are individuals with visual dysfunction (VD) who have the same test score as asymptomatic “N” individuals. This phenomenon of somewhat below average visual skills causing symptoms in some and no symptoms in others poses a diagnostic challenge. A given test score, let’s say 28th percentile, is meaningful only if the visual skill in question is not meeting the patient’s visual demands and symptoms ensue. It is therefore important that symptoms play a key role in a rating scale used to determine whether or not visual dysfunction is present.

**Symptomology**

Optometrists who test visual function in depth and routinely provide VT are convinced that there is a direct cause and effect relationship between deficient VE and VIP skills and symptoms. Research showing this relationship has, with some notable exceptions, defined deficient VE and VIP, a priori, as a test score or scores 1 standard deviation (16th percentile) or lower below the mean. Two studies that provide strong evidence that convergence insufficiency (CI) causes symptoms used a near prism bar positive fusional vergence break (PFV) of 1SD below the mean or worse as a criterion for CI. Another CI criterion was NPC break of 2SD below the mean or worse. A third study using these 2 criteria showed that CI related symptoms resolved as the near PFV break and NPC break improved via office based VT. While these studies provide firm evidence that 16th percentile is the correct criterion, two of them found normal (normal per the criterion used) subjects who were symptomatic. For example, 6% of CI subjects

---

**Figure 1** Four possible diagnostic outcomes in testing visual skills. In the F/VD and P/N outcome, the test result accurately represents the true state of the visual skill.
reported that their eyes felt tired/uncomfortable “most of the time” when reading or studying. For the normal group (n=218) 2% had this symptom most of the time and 28% had it some of the time. Another clear-cut CI symptom, diplopia/words splitting into two, was reported as 5% for CI and 1% for normals on a “most of the time basis”. On a some-of-the-time basis, 17% of CI and 14% of normals had this diplopia symptom.

This CI study showed that there were subjects who were normal based on scores above a predetermined pass/fail criterion yet had the same symptoms as abnormal, CI subjects. Were they symptomatic because they reported symptoms they do not have? Were the symptoms secondary to something other than a visual problem? Or, were there subjects incorrectly classified as normal because the pass/fail criterion was too low for those subjects and the subjects had genuine symptomatic visual dysfunction? I believe the latter is a distinct possibility.

Accommodative facility (+/- 2.00 flippers at 40cm), negative fusional vergence, positive relative accommodation, saccades/pursuits, Visual-Spatial and Visual-Motor Integration have all been studied and linked to symptoms when deficient is defined as 1 SD below the mean. The symptoms of difficulty coming to attention and sustaining attention were linked to inefficient visual skills in a group of twenty school children. The inefficient visual skills diagnosis was based on accommodative and vergence test scores that were one or more SD below the mean. A study of binocular accommodative facility, as measured by the amplitude-scaled method, and symptoms, concluded that fail is 10 cycles per minute (CPM) or less. For children (age 8-16 years) 10 CPM is 0.69 SD below the mean, which converts to the 25th percentile. For adults (age 23-37 years) the 10 CPM criterion converts to the 11th percentile.

Although 1 SD below the mean dominates the literature, there are differing study results, viewpoints, and conclusions. Vergence facility and its relationship to symptoms were investigated using the 3BI/12BO flip prism test. It was concluded that 15cpm or lower is a fail. 15cpm is 0.39 SD below the mean and converts to the 35th percentile. This study by Gall, Wick, and Bedell reconsidered the traditional 1 SD below the mean criterion based on the presence of symptomatic subjects with scores less than 1 SD below the mean. As expected, an upward shift in the failure criterion increased sensitivity but decreased specificity. That is, the new criterion captured more of the symptomatics as deficient vergence facility but it failed more of the asymptomatics. In a recent study of 3BI/12BO vergence facility and symptoms, Gall and Wick revised their pass/fail criterion down to one standard deviation below the mean (≤ 12 CPM). It resulted in a very good specificity (0.88) and mediocre (0.50) sensitivity. Another study of vergence function with prism bar ranges and modified Thorington

![Figure 2. Hypothetical distribution of individuals with visual dysfunction (VD) on a percentile rank scale with no-visual dysfunction (N) individuals.](image-url)
phorias in young children concluded that 0.50 SD below the mean is the lower limit of normal.²⁵

Nearpoint of Convergence (NPC) has been researched twice for normative values in large scale studies almost ten years apart.²⁶²⁷ Both resulted in a conclusion that NPC break of 6cm or higher is a fail and diagnostic of CI.²⁶²⁷ Moreover, both integrated symptom status with the normative data to reach the recommended NPC break value. In the first study, 6cm is a full standard deviation below the mean and converts to 15th percentile for kindergartners.²⁶ For 3rd and 6th graders, however, 6cm is 21st and 31st percentile respectively.²⁶ In the second study, 6cm NPC break is 10th percentile for 6 year olds and 16th percentile for 8 year old children.²⁷ For 7 year olds it is 32nd percentile and for 9 year olds, it converts to 20th percentile.²⁷ Thus, NPC break can be abnormal and associated with symptoms when it is less than a full SD below the mean for some, but not all children.

The relationship between the symptoms (difficulty keeping place while reading, etc.) expected in Oculomotor Dysfunction (OMD) and DEM performance was researched in a clinical setting.²⁸ OMD was defined as a DEM Ratio score of 30th percentile or worse (greater than 0.5 SD below the mean). Symptom status was determined prior to testing with a yes/no symptom checklist. DEM Ratio had a sensitivity of 90% in that 90% of the 48 symptomatic subjects had OMD per failed DEM Ratio. DEM Ratio, per the predetermined pass/fail cutoff of 30th percentile, specified 80% of asymptomatic subjects as normal (no OMD). If a lower pass-fail cutoff of 16th percentile was used, sensitivity would have dropped to 54%. Specificity would have improved to 92%. If a DEM Ratio score of 7th percentile was used in accord with the LVRP-CPG²⁶ and a specific recommendation by Richman and Garzia for DEM²⁹ sensitivity drops to 23% while specificity reaches the very high level of 97%. Stated another way, a 7th percentile criterion would result in 37 of 48 schoolchildren with symptoms expected in OMD diagnosed as no OMD because their DEM Ratio score was above the 7th percentile. VT for OMD would be withheld.

In a Visual-Motor Integration (VMI) study, children were classified as deficient VMI if the Beery Developmental Test of Visual Motor Integration resulted in a performance score of 30th percentile or lower.³⁰ There were 46 subjects with deficient VMI and 41 (89%) of them had symptoms (poor penmanship, difficulty placing written items on paper, etc.) that are believed to be consequences of deficient VMI.³⁰ Of note, 96% of subjects deemed to have a severe VMI deficit, 16th percentile or lower, were symptomatic. 83% of subjects with a moderate deficit (moderate = 17-30th percentile) had symptoms.³⁰

VMI as measured by the Wold Sentence Copy Test was recently studied by Maples.³¹ His study showed that a failure criterion of >0.5 SD below the mean (30th percentile or worse) was the most appropriate for predicting academic trouble.³¹ A normative study of the Rosner Test or Visual Analysis skills resulted in updated norms using percentile rank.¹⁰ The authors empirically concluded that a score below 25th percentile on the TVAS is likely to warrant intervention.¹⁰ VMI (Developmental test of VMI), saccades (DEM Ratio), and rapid automatized naming (RAN, DEM Vertical) were measured in a group (n=33) of schoolchildren with visually based symptoms (the experimental group) and a control group (n=31) who had zero symptoms.³² The symptom status of the study subjects was based on their classroom teacher completing a 12 item symptom checklist at the end of the school year.³² The symptomatic subjects performed worse on the measures of VMI, saccades, and RAN than the asymptomatic subjects and the difference was statistically significant.³² Of note, the symptomatic group had mean performance scores that were at or near the 40th percentile. The asymptomatic subjects had group means that were at or near the 57th percentile. This study provides evidence that visual skills that are only mildly depressed to the 40th percentile can have consequences (i.e. symptoms).³²

Solan and Ficarra studied a group of 51 reading disabled (RD) children with average intelligence and determined the group mean percentile rank with four VIP tests; Developmental Test of VMI, KABC Visual Closure, PMA Spatial Relations, and PMA Perpetual Speed.³³ The group means, 24th percentile, 30th percentile, 38th percentile, and 40th percentile respectively, differed significantly (2 tail t-test, p<.02) from the established means for the tests. All 4 of these group means are well above the 16th percentile.

A review of the literature raises the distinct possibility that symptomatic visual dysfunction can occur in individuals whose visual skill test scores are above the 16th percentile. The final determination that an individual has a genuine visual dysfunction that warrants a VT program, however, goes beyond the percentile rank score of the tests. It takes place at the case analysis phase of an evaluation in which
the optometrist reviews the case history, test results, behavioral observations, academic performance and the complex interactions of all.

Case Analysis
A successful case history elicits all symptoms and concerns that may be due to visual dysfunction. A thorough VT work-up includes sufficient testing of all possible deficient visual skills and relevant observations made during testing. Once the case history and testing are completed, a formal case analysis commences and the examiner concludes that certain symptoms/concerns are rooted in visual dysfunction and uses this conclusion to prescribe or forbear VT. Figure 3 shows four outcomes (two of which are flawed) of the analysis of symptoms and test results. The outcome in box 1 is a straightforward VT case and box 4 is a no-VT case. The outcomes in boxes 2 and 3 are mismatches due to flaws in case history or testing. A discussion follows.

Test Results Low/No Symptoms (Box 2 in Fig. 3)
This result will occur if the pass/fail criterion is set too high. The test scores are correct and represent normal function but incorrectly classified as low instead of normal. In the case of correctly classified low test results, the absence of reported symptoms could be due to a reticent symptomatic patient. I have encountered this problem in children who state that everything is normal as a means to avoid getting glasses. Another possibility is that the symptoms are present but not elicited by the case history. The examiner may not have asked the correct questions to the appropriate individuals. For example, a tutor who listens to a child read aloud on a regular basis could correctly answer a question about loss of place while reading, while a classroom teacher, parent, or the child himself may not.

Another cause of the low test scores/no symptoms mismatch in box 2 of Figure 3 is artificially low test scores with an asymptomatic patient. The visual skill(s) evaluated is truly normal but the test score is an incorrect under-representation of it. Poor understanding of the instructions, poor compliance with the instructions, poor effort, inconsistent attention, and impulsivity are behaviors that can cause a low test score as opposed to visual dysfunction. This problem of sub-optimal test taking behavior is not entirely the responsibility of the patient. It is incumbent upon the examiner to give instructions and engage the patient in such a way that the visual skill is accurately represented in the test score.

Normal Test Results/Visual Symptoms Reported (Box 3 in Fig. 3)
This mismatch could be a poor sensitivity ‘P/VD’ error (Fig 1). The pass/fail criterion is set too low causing symptomatic abnormals to be incorrectly classified as normal. Other patients who report genuine visually based symptoms may have normal test results because of insufficient testing. The all too common example of this error occurs when the only tests administered are for ametropia and eye disease and the results are normal. The eye doctor concludes vision is normal and the symptoms are ignored. This mismatch could be due to errors on the other side of the equation; the test results are correctly classified as normal but the reported symptoms do not exist. A patient may report or exaggerate symptoms they truly do not have.

Another important consideration when test results are normal but symptoms are present is the fact that symptoms may be present for reasons other than visual dysfunction. I often encounter this alternative etiology dilemma when determining the cause of the symptom of frequent decoding errors. The LRVP-CPG lists this symptom as a possible consequence of visual analysis skill deficiency and visual-verbal integration deficiency. Frequent decoding errors, however, are a direct consequence of Dyslexia. Dyslexia, using the definition by Griffin, Christenson and colleagues, is not a visual problem. Indeed, frequent decoding (and encoding) errors in the presence of normal visual skills adds crede to a diagnosis of Dyslexia.

It is clear that case analysis is a complex paradigm that is prone to incorrect conclusions if there are errors in the process of gathering information and clinical data. Symptom errors can be minimized with a thorough case history that uses standardized, validated questionnaires and involves all individuals

<table>
<thead>
<tr>
<th>Visual skill test results low, predict symptoms reported</th>
<th>Corresponding visual symptoms reported</th>
<th>No visual symptoms reported</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Prescribe VT with confidence</td>
<td>2. Extended case analysis required</td>
<td></td>
</tr>
<tr>
<td>Visual skill test results normal, predict no symptoms</td>
<td>3. Extended case analysis required</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. No VT indicated</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3 Case analysis. Reconciliation of visual skill test results and symptoms/concerns from case history
who could observe symptoms. Alternative etiologies for symptoms can be identified by administering tests such as the Dyslexia Determination test or reviewing reports from other specialists who have evaluated the patient.

Among all the potential flaws in the case analysis paradigm, I believe that identifying and interpreting artificially low test scores is the most common and most challenging. It is interesting to note that visual skill testing errors almost always result in false low test scores, not false high. A VT evaluation is therefore biased toward diagnosing visual dysfunction when there is none, and by extension, prescribing VT when it is not needed. One way to mitigate this bias is to set the pass/fail criterion very low as depicted by C₁ of Figure 2.

In my opinion, there is a better way to mitigate the bias toward artificially low test scores. Clinicians who administer VT evaluations should develop and employ a sophisticated set of communication skills and affect that draw out the patient’s best effort, attention and understanding so that the visual skill is correctly represented in the test score. A clinician should also be adept at observing behavior and open to the possibility that, despite the clinician’s best effort, a test score is low because of sub-optimal test taking behavior, not visual skill dysfunction. A clinician confident in her/his ability to administer visual skill tests and judge the veracity of their outcome could follow a recommendation made by Solan and Suchoff that “Any score below the 50th percentile is suspect, especially when a child’s IQ is average or better, since the criterion is mastery of the task”.

Rating Scales Utilized Outside of Optometry

The educators, psychologists, and other clinicians (speech/language, occupational therapy, etc.) who routinely administer standardized, norm referenced, academic, developmental, or performance tests do not utilize a uniform rating scale as shown in Table 2.

Moreover, none of the resources listed provide a rationale for the cutoff values chosen other than a statistical symmetry with standard deviation or the derived score. It appears that the great majority of rating scales in place outside of optometry are not supported by extensive research. Martin, in her chapter on interpretation of TVPS (3rd edition) test scores, acknowledges the arbitrary basis of rating scale systems. She writes:

Functional classifications such as “average”, “below average”, “above average”, etc. may be defined by standard accepted practice within professional fields or by test authors. While such classifications provide a clinically useful “shorthand” of sorts, examiners are cautioned that there is no universally accepted rule as to how many units (i.e., standard deviations, standard scores, percentile ranks) define each functional category.

<table>
<thead>
<tr>
<th>Table 2. Non-optometric rating scales</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CRITERIA</strong></td>
</tr>
<tr>
<td>1. Below 50th percentile is inferior performance</td>
</tr>
<tr>
<td>2. Reading level below 37th percentile is mild Dyslexic</td>
</tr>
<tr>
<td>3. Reading or spelling below 31st percentile is “delayed”</td>
</tr>
<tr>
<td>4. 1st-25th percentile is “below normal range” and “weakness”</td>
</tr>
<tr>
<td>5. Reading level of 25th percentile or lower is “poor reading”</td>
</tr>
<tr>
<td>6. 9th-24th percentile is below average</td>
</tr>
<tr>
<td>7. 9th-16th percentile is below average</td>
</tr>
<tr>
<td>8. 2nd-15th percentile is below average</td>
</tr>
<tr>
<td>9. 2nd-12th percentile is “low” 13th-87th percentile is “average”</td>
</tr>
<tr>
<td>10. 9th-24th percentile is “low average” 2nd-8th percentile is “borderline” 1st percentile or lower is “extremely low” or “delayed”</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>SOURCE</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Professional fields or test authors.</td>
</tr>
<tr>
<td>Martin’s chapter on interpretation of TVPS (3rd edition) test scores.</td>
</tr>
</tbody>
</table>

Optometry & Vision Development
Martin provides three examples of categorization schemes and comments:

It is important for the examiner to realize that these classification schemes represent a traditional practice and are not set in stone (statistically speaking).³⁵

Although there is a wide range of reported rating scales, criteria 6 and 10 in Table 2 are the most common ones I encounter. Criterion 6 is important because parents of schoolchildren see this scale as it is applied to government-mandated academic achievement tests. For schools using this criterion, consider a parent who is informed that his/her child has below average reading because of 20th percentile reading achievement test scores. The parent consults an optometrist who administers a battery of VIP tests to their child and the results cluster at/near the 20th percentile. This level of VIP function is well above the LRVP-CPG 7th percentile-is-anomalous criterion and higher than the LRVP-CPG 8th-16th percentile is “suspicious and perhaps clinically relevant.” Thus, VIP could be rated as satisfactory and VIP VT would be withheld. An astute parent could ask why 20th percentile is below average in reading but considered a satisfactory level of VIP function.

Criterion 10 in Table 2 is from the ubiquitous Wechslser Intelligence Scale for Children (WISC). The scale is applied to many other tests and one textbook refers to it as a “commonly accepted classification of ability levels”.³³ A noteworthy feature of this scale is its use of the label “low-average” for the range of scores from the 9th-24th percentile (80-89 standard score.) The term low-average is ambiguous and does not clearly specify if a concern is present that should trigger action. I perused the charts of 15 VT patients in my office that had a psychoeducational report that utilized the WISC scale with its low-average classification.⁵¹ All 15 reports were from different examiners. Eleven of the 15 interpreted a low-average score in 9-24 percentile range as a concern that triggered action (referral for VT evaluation, reading therapy, speech/language therapy, etc.). The other four interpreted low-average as a no-concern test result. One child in this group was tested by an educational psychologist on the east coast. Her score on the Beery VMI was 14th percentile, WISC Block Design was 16th percentile (scaled score=7) and WISC Coding was 25th percentile (scaled score=8). All three of these tests were rated “low-average”. The examiner concluded that “Visuomotor function is satisfactory” despite the fact this 3rd grade girl had poor penmanship, difficulty copying and difficulty aligning written math problems. Her family moved to Pasadena, CA and another educational psychologist reviewed the previous report. She made a prompt referral for a VT evaluation because of these “low-average” scores.

Proposed Alternative To The LRVP Rating Scale

The scale I propose and presently use is shown in Table 3. It is a blend of recommendations made by

<table>
<thead>
<tr>
<th>Symptom Status</th>
<th>Test Result</th>
<th>Cognitive Level</th>
<th>Cluster</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symptomatic</td>
<td>1-30th %ile</td>
<td>deficient</td>
<td>Yes</td>
<td>Deficient VT Indicated</td>
</tr>
<tr>
<td></td>
<td>31st-39th</td>
<td>low-average</td>
<td>No</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td>40th-59th</td>
<td>average</td>
<td>High</td>
<td>Deficient VT Indicated</td>
</tr>
<tr>
<td></td>
<td>&gt;60th</td>
<td>above average</td>
<td>Normal or Low</td>
<td></td>
</tr>
<tr>
<td>Asymptomatic</td>
<td>&gt;17th</td>
<td>Low</td>
<td>No</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td>1-16th</td>
<td>Normal or High</td>
<td>Yes</td>
<td>Deficient</td>
</tr>
</tbody>
</table>
Solan/Suchoff, Borsting, Groffman, Scheiman/Galloway, and Mozlin. Solan and Suchoff make the point that 1SD below the mean is a convention that comes out of research that seeks to label findings as undoubtedly poor. Their specific recommendation is that scores below the 31st percentile “…be considered poor and represent a real dysfunction…” 7

Borsting was the first to link symptoms to a rating scale. He recommends the following scale for rating the performance scores generated by VIP tests.

Performance that is one standard deviation below the mean (16th percentile) or worse is abnormal and, in most cases, should be managed with an appropriate intervention program. Performance that is between the 16th and 35th percentiles is considered weak or suspect and may represent a problem for the child. A confirmation with signs and symptoms in the case history will indicate if an anomaly is present.

Scheiman and Galloway specify that below 16th percentile is significantly low. They acknowledge that the range of scores from 16th-30th percentile are “suspicious findings and may be clinically significant when compared with the child’s signs and symptoms and overall level of functioning.” Groffman/Solan rate visual skills as “a little below average” if they are 23-39th percentile (4th stanine). 11th-22nd percentile is below and 4th-10th, 0-3rd are low and very low respectively.

Groffman, Solan/Suchoff, and Mozlin discuss the importance of behavioral observations and supramodal processing in the assessment process. Behavioral observations can and should be used to conclude that visual dysfunction is present despite a normal test score. For example, a patient who leans backward, ceases blinking, and experiences visual discomfort during the NPC test, may have a convergence problem even though the NPC break and recovery equal the expected score of 5cm/7cm. Behavioral observations can drive decision making in the other direction. A patient taking a visual skill test with an impulsive approach or inadequate attention may score artificially low as discussed previously. A pattern of deficiencies or strengths in one of the two modes of supramodal processing (simultaneous and successive) can aid interpretation of a low test score. Consider a patient who scores low on a visual memory test that requires simultaneous processing and scores high on visual memory tests that require successive processing. The low score is meaningful if there is a cluster of low scores on other tests that involve simultaneous processing. The utilization of a cluster of test scores within a diagnostic category is also applicable. A binocular cross cylinder finding that is slightly low (eg. +0.25) is less meaningful if MEM, NRA, and plus side of accommodative facility test are normal and vice versa.

The paradigm for rating a VE or VIP visual skill test score from a VT work-up, shown in Table 3, is predicated on four factors. These factors include symptom status (“S” or “A”), test result, cognitive level and cluster. The end result of the paradigm is to diagnose visual dysfunction that warrants VT or to conclude that visual function is normal.

**Symptom Status**

Symptom status, garnered from a thorough case history, is the first step and dominant factor in the paradigm. Symptomatic means at least one definite symptom is present that is a likely consequence if the visual skill in question is deficient. A symptomatic patient can score as high as 30th percentile on the test and have visual skill dysfunction requiring VT regardless of cognitive level or cluster status. An asymptomatic patient does not have visual symptoms that could be associated with visual skill. For these, a diagnosis of visual dysfunction is contingent upon a poor score (≤ 16th percentile), a cognitive level of normal or higher, and a cluster of low scores. It should immediately be pointed out however, that this clinical scenario is rare in my experience. In other words, a patient with a legitimately poor visual skill test score within a cluster of low scores who has satisfactory intelligence is almost always symptomatic.

The key role of symptom status in the diagnosis of visual dysfunction exposes a flaw in the paradigm. The body of scientific evidence supporting a causal relationship between visual skill dysfunction and symptoms is incomplete. Within this void, examiners impart judgement as to the basis of a reported concern. Their judgment is an expression of their clinical experience and model of vision. For example, the concerns, “answers questions before they have been completely asked” and “difficulty with money concepts, making change,” do not intuitively appear connected to visual function nor is there research linking them to vision but they have been reported as such. An examiner whose model of vision connects visual function to a broad range of behavioral concerns is apt to label a
patient as symptomatic and bias the assessment toward diagnosing visual skill dysfunction/prescribing VT.

**Test Result**

For symptomatic patients, the percentile rank score from the visual skill test is rated as deficient (1-30th percentile), low average (31-39th percentile), average (40-69th percentile), or above average as shown in Table 3. The upper end of deficient, 30th percentile, is the criterion recommended by Solan/Suchoff. The low-average and average categories allow cognitive level and cluster to influence the final rating as explained below. The above average category (> 60th percentile) always converts to normal function. An extended case analysis would be required to ascertain the cause of the symptom.

For asymptomatic patients, there are only two test result categories. If the test result is >17th percentile, then the visual skill sampled by that test is normal. If the test result below 17th percentile may represent a deficient visual skill when cognitive level and cluster factors are analyzed as explained below.

**Cognitive Level**

Cognitive level influences diagnosis in two clinical scenarios. The first, and perhaps the most common, is a case of a symptomatic patient with a visual skill test score in the average range. If this symptomatic patient has a high cognitive level, especially if it is high because of strengths in verbal intelligence, the patient has visual dysfunction and VT is indicated. It is preferable that the cognitive level has been formally ascertained by an educational psychologist or neuropsychologist. The second scenario is an asymptomatic patient for whom the visual skill test resulted in a low score of ≤16th percentile. If cognitive level is low, the score does not equate with visual dysfunction. If normal or high, the asymptomatic patient may have visual dysfunction if a cluster of like visual skill tests result in low scores of ≤16th percentile.

**Cluster**

Cluster refers to the existence of at least one additional low-test score in the same diagnostic category or supramodal (simultaneous versus successive) processing. For example, a low NPC score and low near PFV score are a cluster for deficient convergence. A low score on the TVPS Visual Closure subtest coupled with a low score on the Divided Form Board would be a cluster depicting a weakness in simultaneous supramodal processing. Cluster has the least impact in this paradigm. An asymptomatic patient with a poor test score requires a cluster to be diagnosed with visual dysfunction. Cluster can also aid interpretation of borderline scores that are in the low-average range of 31st - 39th percentile for a symptomatic patient. Solid average scores on companion visual skill tests, i.e. no cluster, may justify ignoring the borderline score. Not shown on this paradigm is the utilization of cluster to establish severity of visual skill dysfunction. If five VIP tests requiring sequential processing are administered and two have a low score the sequential processing weakness is less severe than if all five were low.

**Conclusion**

Optometrists specializing in VT are charged with differentiating those with visual skill dysfunction requiring treatment from those with satisfactory visual function. An evidence-based visual skill test rating scale that consistently augments case analysis would be a significant clinical asset. It is my opinion and clinical impression that symptom status can and should influence the rating of a visual skill test score. A rating scale tethered to symptoms is patient-centered as opposed to statistics centered. To obtain an undisputed evidence-based status, a symptom-driven rating scale will require significant supporting research showing a causal link between symptoms and deficient visual skills. The symptoms in the link should be specific as possible and gathered with case history methods that are valid and repeatable as has been done with the Convergence Insufficiency and Reading Study symptom survey. Moreover the visual skill tests utilized in this type of research should be standardized to delineate not only the level at which symptoms are probable but possible.

The rating scale I have proposed is not entirely original. It is a synthesis of previously published rating scales, criteria, and case analysis strategies. I acknowledge that it, like all other rating scales inside and outside of optometry, have arbitrary cut-off points. Clearly, there are other complex relationships, germane to a rating scale, which this essay did not explore. For example, if accommodative and vergence skills are satisfactory, VIP function may have to be lower to cause symptoms than if accommodation/vergence were deficient. Another possible influence on a rating scale is age/grade. Research has shown that the VIP skills measured by the Divided Form
Board test have a higher correlation with reading in kindergarten versus second grade.\textsuperscript{61} Thus, a score of 30th percentile on this test from a kindergartner may be a significant concern while the same score from a 2nd grader causes no concern. The fact that the cutoff for NPC break performance scores varies by grade is another example.\textsuperscript{26,27} Another layer of complexity is the possibility that each visual skill may have its own scale and/or each test may require a unique scale. This article did not explore the influence of a flawed test on a rating scale. Clearly, a visual skill test that is unreliable, invalid, or based on weak normative data can provide results that are not representative of the visual skill under scrutiny and cause the rating to be incorrect. Groffman provides a thorough discussion of the characteristics that make for an appropriate visual skill test.\textsuperscript{62}

The main point of this essay is to express my opinion that the LRVP-CPG criterion is too low. It is too low because there are individuals who test normal per the LRVP-CPG criterion but still have symptomatic abnormal visual function. If this criterion is widely adopted at the schools and colleges of optometry and clinical practice, I believe there will be too many individuals diagnosed with normal visual function when in fact they have significant visual dysfunction requiring therapy. I am also concerned that the LRVP-CPG scale could limit access to VT in another way. A third party could use the LRVP-CPG scale and determine VT is indicated only when visual skill test scores are below the 7th percentile. (The LRVP-CPG is accessible to the public through the AOA website.) Moreover, a VT program could be deemed finished once visual skills rise above 7th percentile but are below 31st percentile – that could be a premature discontinuation of treatment services. I suggest that the AOA Consensus Panel on Care of the Patient with Learning Related Vision Problems reconsider their criterion for an anomalous clinically significant VIP test result if and when a revised CPG is published.

Acknowledgement: Drs. Hana Khoo and Tony Carnevali provided numerous helpful corrections, deletions, and suggestions during the preparation of this manuscript.

References:


