

The Beery-Buktenica Developmental Test of Visual-Motor Integration:  
An Overview, Evaluation, and Critique



Chapman University

The Beery-Buktenica Developmental Test of Visual-Motor Integration:  
An Overview, Evaluation, and Critique  
Description

The subject of this test critique is the Beery-Buktenica Developmental Test of Visual-Motor Integration (Beery VMI) Sixth Edition. The Beery VMI was authored by Keith E. Beery after he noticed that children's abilities to copy geometric forms correlated with their academic achievement. Natasha A. Beery and Norman A. Buktenica, who designed four of the geometric forms used in the assessment, are coauthors. The Beery VMI was first published in 1967 as The VMI, and the sixth edition was published by Pearson in 2010.

The Beery VMI is a norm-referenced assessment that involves copying a sequence of geometric forms. The Beery VMI consists of the Visual-Motor Integration (VMI) scale, which is the core task, as well as two additional tasks, the Visual Perception (VP) and Motor Coordination (MC) subtests (Beery & Beery, 2010). The VP subtest evaluates clients' visual abilities without the integration of fine motor skills. Examinees point to items and pictures from the manual to identify gross differences and objects, and the administrator records all responses. The MC task assesses fine motor skills independently of visual perceptual skills by requiring clients to connect dots and draw within bordering lines.

Content and Use

The Beery VMI presents the client with 30 drawings, increasing in difficulty, to copy using a pencil. It was designed to assess individuals' abilities to integrate their visual and motor skills, and is used for children, adolescents, and adults through 100 years old. Because administration is quick (approximately five to fifteen minutes), it is commonly used in educational, medical, and research settings. The developers state that the Beery VMI is useful as

an initial measure in an assessment battery because it helps establish rapport with children (Beery & Beery, 2010).

The \$138.00 Beery VMI starter kit includes the 210 page 6<sup>th</sup> Edition manual. While it is not always clear and is rarely concise, the manual does include teaching suggestions. Also included are 25 VP forms, 25 full forms, 25 short forms, and 25 MC forms.

The instructions for administering the Beery VMI are easy to understand. They could be improved by providing tabs in frequently used sections and by writing them in bold font so they are easier to see on the page. Although the directions have been developed so that clinicians can read the same directions to children and adults, examiners might sound childish if they read them as written to adult examinees.

To complete the assessment, the respondent copies geometric drawings on a form. However, there are no specific guidelines in the manual regarding the process examiners use to determine a child's functional age, and thus, their start point. Scoring is done directly on the forms, which is convenient for administrators. In the sixth edition, the adult and child forms are combined into one, which is convenient for administrators who work with clients of differing ages (Beery & Beery, 2010).

The VMI can be implemented as part of the Response to Intervention (RTI) model, as it is easily administered to school classrooms for screening purposes, and to smaller groups and individuals who are in need of an intervention. In addition to serving as a screener, the Beery VMI is used to obtain the services that clients need, to measure the effectiveness of interventions, and to advance research. A strength of the Beery VMI is that the assessment is virtually culture-free, and is useful with individuals of diverse environmental, educational, and linguistic backgrounds because it uses geometric forms instead of letter or numeric forms. This

reduces cultural bias in the test that may result from individuals from different backgrounds having varying degrees of exposure to alphabets and numbers.

#### Standardization Sample and Norms

The developers of the Beery VMI ensured that the assessment was normed on a sample that represents the diversity of the United States. From late 2009 to early 2010, the VMI was re-normed on children from four major census regions across the country. The norming samples are representative of the 2010 census not only in regions, but also in gender, ethnicity, parent education level, residence, and age. The adult norms are based on data from late 2005 through early 2006 involving 1,021 adults from 27 states across all major regions in the United States. The sample was reasonably representative of the 2000 census in terms of gender, ethnicity, residence, region, and age. The demographic characteristics for both children and adults are presented in a clear format in the manual.

Researchers should strive to use norming samples that are representative of the populations for which the test will apply. Large sample sizes yield more generalizable results, and the sample size for the Beery VMI is acceptable. However, the method of recruiting is not representative of an appropriate standardization technique. School psychologists from different regions in the United States were randomly selected, and asked about their school populations. From these schools, norming participants were chosen based on need according to the 2010 U.S. Census. All of the children in the sample classes, including those with disabilities, were included in the norming data. Although the data represent the population, the assessment could have bias because entire classes were used instead of using students from different classes, schools, and districts.

#### Scores and Interpretation

Scoring is a relatively easy process for examiners, although it requires careful analysis because measurements and degrees of angles determine scores on some items. Scoring guidelines are available in the Beery VMI manual. Examples of incorrect and correct responses are provided for each item on the VMI. This is a helpful reference for scorers.

Examinees earn one point for each correctly copied item, and ceiling is reached when three consecutive items are incorrect. The manual suggests that if scorers are in doubt, they should “score the item as passed” (Beery & Beery, 2010, p.28 ). To obtain the examinee’s raw score, the examiner subtracts the incorrect items completed prior to the ceiling from the ceiling. The manual provides a clear description of obtaining the raw score, as well as a helpful example for users of the Beery VMI.

The manual describes the utility of age norms for explaining to parents their child’s current developmental level (Beery & Beery, 2010, p. 27). It gives a clear definition of age norms- “the age at which about 50% of children meet the developmental criteria for a given form”. On page 177, Beery explains that age equivalents, as well as grade equivalents, “should be used judiciously, if at all”. Although the section does not further describe the inaccuracies of age and grade level equivalents, it is beneficial to warn against over-interpretation of them. Many examiners are unfamiliar with the problems of using age and grade equivalents and may over-interpret them in reports.

The author continues to explain that percentile ranks and standard scores are more reliable and valid measures. Percentile ranks are especially useful for identifying outliers and extreme scores. Since scores gather around the mean, they are not a useful measure for individuals with relatively average scores because small differences are exaggerated. However,

standard scores are more useful. They lack precision, but are used to interpret all tests the same way, which makes them easy to understand.

Another weakness of the Beery VMI is that there is no mention of confidence intervals in this section of the manual. It is important for users to remember that the scores they obtain are not truths; all scores include some degree of error. Without an explanation of confidence intervals and their significance, users may be prone to over-interpretation of standard scores.

### Psychometric Properties

#### Reliability

The Beery VMI's reliability was measured in terms of item content, re-administration scores, and inter-rater scoring. Since the VMI is used for clinical and educational decisions about individuals, it should have a reliability of .90 or higher. Norming results have correlated almost perfectly since the test's inception. Thus, the sixth edition cross-validates norms instead of repeating studies done with large norming populations.

Content sampling. Rasch-Wright analysis is used to assess the extent to which the test items align with the author's test construct, how well they are separated from one another, and how well they differentiate among individuals (Beery & Beery, p. 103). Next, the manual explains that Rasch-Wright analyses are most appropriate for power tests (tests in which the test taker is not pressured to work quickly) such as the VMI. However, the manual fails to present a definition of the Rasch-Wright analysis and its utility in measuring reliability.

The results indicate mean content reliability correlations between .93 and .96 and mean person reliability correlations between .83 and .84 for the Beery VMI and its subtests. Users of the Beery VMI can assume that these results are accurate because a random sample of 50 individuals from each age group of the fourth edition norming data was used. However, these

correlations do not meet the criteria of .90 or higher, and thus the Beery VMI may not meet reliability standards for use with individuals.

Internal consistency. The manual explains that test items should be homogenous throughout the assessment. It then explains in simple terms the Spearman-Brown method and lists the results of an even-odd split. The correlation between halves was found to be .95, which is strong. For educational and clinical purposes, a correlation of at least .90 is necessary.

Another method of assessing internal consistency used is coefficient alpha. The manual explains that it splits and correlates items in every possible way and that it is a valid measurement of internal consistency because it increases with intercorrelations among test items. Chronbach's alpha for the Beery VMI was found to be .96, which is very strong. Overall, this section of the manual provides both clear explanations and appropriate data.

Standard error of measurement. The Beery VMI manual explains that there is error in every measurement because statistics is based on probabilities. It also explains how to use the standard error of measurement (SEM) to calculate a 68% confidence interval. However, it fails to explain what the SEM is, and its important in test construction. The manual instructs users to "interpret these SEMs in the usual manner" (p. 107). Unfortunately, many individuals who use the Beery VMI will not know the correct way to interpret the SEM, and since the manual does not provide a formula for confidence intervals, users who are not aware of the formula may find this section confusing. They will also be unable to create a confidence band at 90% or 95%, and will have to use the 68% band provided in the book. Furthermore, the manual should explain that scores should be interpreted in relation to the amount of error in a measure. Test administrators who do not have a background knowledge in statistics may not understand the importance of this section, and that may lead to over-interpretation of results. Lastly, the manual does not express

the difficulties in interpreting results when a confidence interval is larger than 30 units. However, the highest SEM for the Beery VMI is 6, so that problem is avoided in this instance.

Time sampling. To assess the extent to which students' scores remain consistent, 142 children between the ages of five and twelve were given the Beery VMI in January of 2010. Time between the initial administration and the retest was 14 days on average. The test-retest coefficient for the VMI is .88, .84 for the VP, and .85 for the MC subtest, all of which exceed guidelines for retesting between one and two weeks.

There are several weaknesses of the test-retest data. For example, although the manual explains that the children were from a "regular public school classroom" that has a proportionate number of students with disabilities, no additional demographics are given (Beery & Beery, 2010, 107). Preferably, the manual would state the proportion of sexes and ethnicities, and they would match those of the original administration. Second, 1737 children were used to ascertain norms for the Beery VMI. However, the retest group only utilized 8.2% of that original number. Although a sample size of at least 100 is adequate, the students' ages are not representative of the individuals' for whom the test is normed. To improve the relevance of the test-retest reliability score, the researchers should use a stratified sample in the second test administration as well as the first.

Interscorer reliability. To determine the interscorer reliability for the sixth edition of the Beery VMI, two individuals separately scored 100 tests of a random sample of the children's norming group. Reliability was found to be .93. Interscorer reliability based on 25 adults' scores from the norming sample was found to be .94. These are high reliability coefficients. The manual continues to explain that interscorer reliability has been found to increase with increased preparation of the scorers, regardless of the background of the scorers. This section is easily



understood and explains the utility of interscorer reliability that is found between administrators of differing backgrounds.

### Validity

To achieve validity, a test must first be found to be consistent in its measurements, or reliable, as discussed above. There are also several kinds of validity that must be achieved, and they will be discussed below.

Content validity. The Beery VMI manual describes content validity as “the degree to which the content of a test provides a representative sample of the behaviors the test is designed to assess” (Beery & Beery, 2010, p.111). In simpler terms, this means that constructs on the test should relate to actual classroom requirements; in the case of the Beery VMI, this domain is visual-motor integration: “the degree to which visual perception and finger-hand movements are well coordinated” (Beery & Beery, 2010, p. 13). This can be established through careful selection procedures used in selecting items or tasks for a test.

The manual’s sections covering content validity are not written in a format that facilitates users to ascertain the information they need. First, the authors mention that the test’s construct must be defined and assessed. However, they list background research about visual, motor, and visual-motor development next, instead of explaining the construct. Identifying the specific construct of the Beery VMI is difficult because of the other information that is presented in this section. Additionally, content validity is mentioned in the general validity chapter of the manual. Instead of stating a validity coefficient, the authors direct readers to refer to the item construction and selection procedures found in Chapter V of the manual. They mention that content validity has been assessed for the Beery VMI using the Rasch-Wright analysis method. According to the manual, the content validity of the VMI is “strongly supported”; however, no coefficient is given

(Beery & Beery, 2010, p. 111). This is a serious weakness of the Beery VMI because users do not know if the test measures what it purports to measure.

Concurrent validity. One way to determine an assessment's concurrent validity is to compare the examinee's results on that test to his or her results on another test that measures a similar construct. By comparing the results of the Beery VMI to the results of the Copying subtest of the Developmental Test of Visual Perception (DTVP-2), and the Drawing subtest of the Wide Range Assessment of Visual Motor Abilities (WRAVMA), researchers determined that the concurrent validity of the Beery VMI is moderate. A .52 correlation was found between the VMI and the WRAVMA Drawing, and a .75 correlation was found between the VMI and the DTVP-2 Copying. Researchers also compared the Beery VMI supplemental tests to the DTVP-2 Position in Space and Eye-Hand Coordination subtests, and found correlations of .62 and .65 respectively. The manual cites that the discrepancy is caused by the other tests being "newer, less well-developed geometric form-copying tests" (Beery & Beery, 2010). The Beery VMI manual takes no credit for any flaws in their assessment, which may lead inexperienced users to believe that the Beery VMI is perfectly constructed.

Construct validity. The concept of construct validity is "demonstrated by identifying several constructs thought to underlie test performance, then generating hypotheses based on those constructs, and, finally, verifying the hypotheses by empirical data or logic" (Beery & Beery, 2010). There are several methods for measuring construct validity, and the authors of the Beery VMI utilize many of them. The Beery VMI manual lists and examines the following seven hypotheses:

1. The abilities measured by the Beery VMI are developmental. Thus, it is hypothesized that results from the tests will be related to chronological age.

Since the Beery VMI is designed to measure changes in hand-eye coordination throughout childhood, the Visual Perception and Motor Coordination subtests were designed to be developmental scales. Researchers used the 2010 norming data to correlate chronological age with Beery VMI, Visual Perception, and Motor Coordination scores. The Pearson correlations were .89, .85, and .84, respectively. This validity study was conducted well since all data from the norming sample were used. The test's ability to identify developmental changes indicates strong construct validity.

On the other hand, an experiment done by Iodine, Taub, and Harris (2014) indicates that there was no significant correlation between the VMI and age.

2. The abilities measured by the Beery VMI and its supplemental tests are related to one another because each supplemental test measures a part of that the Beery VMI measures. Thus, it is hypothesized that results from the tests will correlate at least moderately well with one another.

To confirm this hypothesis, researchers used the sixth edition norming data and found that all correlations were significant beyond the .05 level of confidence. This section, however, is difficult to understand. While it is beneficial to establish convergence through correlations with related constructs, the manual must be more clear to be useful for administrators.

3. Each of the Beery VMI supplemental tests measures a part, but not the entirety, of what the Beery VMI measures. Thus, it is hypothesized that there will be evidence the Beery VMI is more demanding than either of the supplemental tests alone.

To test this hypothesis, researchers used the norming data from the fourth through sixth editions of the Beery VMI. They found that children made more correct responses on the Visual Perceptions and Motor Coordination tests than on the Beery VMI, although the time limit on the

subtests is shorter. This suggests that the Beery VMI is indeed measuring the integration aspect of the term visual-motor integration. Thus, the Beery VMI is greater as a whole than the sum of its parts. The utilization of a large sample size is beneficial in validity studies.

4. The abilities measured by the Beery VMI and its supplemental tests are related to at least nonverbal aspects of intelligence. Thus, it is hypothesized that results from the tests will correlate moderately well with nonverbal intelligence test results and less well with verbal intelligence results.

To test this hypothesis, researchers correlated the Revised Wechsler Intelligence Scale for Children (WISC-R), the Beery VMI, and the supplemental tests for 17 children between the ages of six and twelve years who were identified in the fourth edition as having learning disabilities. Correlations are listed in the table below.

	<b>Beery VMI</b>	<b>Visual Perception</b>	<b>Motor Coordination</b>
WISC-R Verbal IQ	0.48	0.43	0.41
WISC-R Performance IQ	0.66	0.58	0.55
WISC-R Full IQ	0.62	0.54	0.51

The reasons why researchers used only individuals with learning disabilities is unclear. This section of the manual lacks details as to how the group was chosen and what the participants' demographics were. A well conducted study should use data from stratified age groups, socioeconomic statuses, ethnicities, etc. Unfortunately, the Beery VMI manual gives no information regarding this information, which would be helpful in supporting the assessment. Another limitation is that this study was conducted with the fourth or an earlier edition of the Beery VMI, so the findings are potentially unrepresentative of the current edition.

Additionally, researchers found that the Beery VMI correlates more strongly with chronological age than with intelligence. Thus, it may be more sensitive for physiological/neuropsychological problems in child development. General intelligence is more mediated by the frontal association areas of the cortex than visual-motor integration, which may be mediated by white matter. By comparing the Beery VMI to intelligence and chronological age, the authors sought to establish convergence, a relationship between the test and related constructs. Correlations should be between .60 and .70 or above; thus the Beery VMI establishes low to moderate levels of criterion validity.

5. The abilities measured by the Beery VMI and its supplemental tests are related to academic achievement. Thus, it is hypothesized that results from the tests will correlate moderately well with academic achievement test results.

Researchers correlated the Beery VMI with the Comprehensive Test of Basic Skills (CTBS) for 44 fourth- and fifth-grade students from regular classrooms and found the following correlations:

	<b>Beery VMI</b>	<b>Visual Perception</b>	<b>Motor Coordination</b>
CTBD Reading Total	0.58	0.31	0.31
CTBD Language Total	0.68	0.2	0.39
CTBD Mathematics Total	0.42	0.21	0.37
CTBD Overall Total	0.63	0.29	0.40

These correlations are low. In order to establish criterion validity, correlations should be between .60 and .70 or above. Furthermore, the manual lacks important information regarding

the study. This correlation is based on data from fourth- and fifth-grade students, but no demographic information is given. This limits the generalizability of the findings.

The information given in this section does not relate to the hypothesis that is being discussed; instead of admitting that the correlation is low and identifying reasons that might cause that, the manual lists several studies conducted by other researchers. Although it is beneficial to have similar results from various sources, the manual does not provide any background information about the studies. This could lead to misunderstanding among some users of the Beery VMI who do not follow up on the information provided.

One such example is the listing of intercultural correlations on page 117 of the manual. It mentions that Beery VMI correlations for reading and mathematics among fifth- and sixth-grade Taiwanese children ranged from .51 to .73. Additionally, correlations have ranged from .42 to .55 for reading among Japanese children ages 11 to 15 years old, and from .65 to .67 for mathematics among the same group.

While it is interesting to note that correlations between the Beery VMI and academic achievement are high among other cultures, this information must be interpreted correctly. This is not to say that the Beery VMI correlates with academic achievement for all other cultures, or even for all Asian cultures. Without further demographic information including sample size and stratification, this information is not adequate to generalize to other cultures.

6. The items of the Beery VMI and its supplemental tests measure similar respective traits and are effective in measuring persons. Thus, the Rasch-Wright item and person separation indices will be high.

The Rasch-Wright indices are high for all age groups. Furthermore, a sample of 50 individuals, selected randomly, from each age group were selected from the fourth edition norming data to provide the sample. This is strong evidence of concept validity.

7. The abilities measured by the Beery VMI are sensitive to certain disabling conditions. Thus, it is hypothesized that Beery VMI results will be lower among populations with those conditions.

A study done by Krab (2008) evinced that the Beery VMI differentiated between groups of either specific or general learning disabilities. It also differentiated between groups without learning disabilities and with general learning disabilities. Significance was found at the .05 level. Children who have Attention Deficit Hyperactivity Disorder (ADHD), brain injuries, developmental coordination disorder (DCD), educable mental retardation, and visual impairments scored less well on the Beery VMI than their peers. However, scores of children with low academic achievement do not differ from scores of children with disabilities.

Another use of the Beery VMI is with children diagnosed with Autism Spectrum Disorder (ASD). Many individuals who have autism have difficulty with visual-motor integration tasks, including writing. Because ASD evaluations must contain assessments of visual-motor control, the Beery VMI has an important use in the medical and research realms.

It is crucial that test authors conduct adequate studies to determine the construct validity of their assessments. An assessment is not valid if it does not measure the concept that it is intended to measure. To strengthen their validity results and add to the current hypotheses of the Beery VMI, researchers could incorporate tests of congruence, and convergence.

Factor analysis. A number of tentative developmental stages have been defined by Polumbinski and others (1986). They found that the acute and oblique angles accounted for the

largest amount of variance (25%) in Beery VMI performance. Factor analytic studies have indicated that visual-motor integration was the key underlying factor for handwriting performance.

The manual provides only brief information regarding factor analysis. Inconveniently, it does not provide readers with a definition of the term, or the authors' method of obtaining the subtests.

Predictive validity. "Letter identification" and "reading readiness", which both rely heavily on visual and visual-motor skills, are predictors of reading difficulties at school entry (Snow et. al, 1998). Thus, the Beery VMI can be useful in predicting achievement. Comparisons of students' pre-kindergarten test scores with their achievement at the end of kindergarten and at the end of first grade indicated that the Beery VMI "best predicted achievement". The assessment also identified high-risk boys in kindergarten who later had difficulties with reading.

However, these studies have not been updated with the current edition of the Beery VMI, so the data are not necessarily representative of current predictive validity. The manual briefly lists several other studies that support the test's predictive validity. These provide users with ideas for using the Beery VMI as a predictor. However, no sample sizes or demographics are listed for readers. Thus, without further investigation into the cited studies, users of the Beery VMI must interpret these findings accordingly, and must refrain from generalizing them to the general population.

### Conclusion

The authors believe that the Beery VMI is "the most widely researched and most valid test of its kind," (Beery & Beery, 2010, p. 1). It displays many strengths, such as easy administration and comprehensive scoring instructions. However, the test has limitations in



crucial domains. First, there is a lack of current validity and reliability research. Much of the psychometric data pertains to previous editions of the Beery VMI and may be invalid according to new conceptualizations of cognitive intelligence, academic achievement, and other related constructs.

Additionally, many of the psychometrics contain weak samples and low correlations. Analysis conducted to determine content sampling reliability did not produce sufficient reliability correlations to appropriate educational and clinical decision making for individuals. Elements of validity, such as congruence with academic achievement, also fail to reach appropriate levels. Additionally, concurrent validity results indicate only low to moderate correlations. These correlations indicate limitations of the Beery VMI's validity.

The subtests of the Beery VMI are particularly weak. The authors mention that, as the technical quality of the supplemental tests is not as high as the VMI, experienced clinicians may find their own clinical methods superior in gauging the examinee's visual perception and motor coordination abilities (Beery & Beery, 2010, p. 100). This statement voids the subtests' utility, and admits that there are other methods, many of which are free of cost, that provide better measurements of clients' visual perception and motor coordination.

Therefore, although the Beery VMI is a well-normed, easy to administer assessment that is commonly used in schools, it should not be used for individual diagnostic purposes. Resulting scores from the test should be analyzed in addition to other forms of a thorough assessment.

## References

- Beery, K.E., Beery, N.A. (2010). *The Beery-Buktenica Developmental Test of Visual-Motor Integration*. Bloomington, MN: Pearson.
- Beery™ VMI 6th Edition Beery-Buktenica Developmental Test of Visual-Motor Integration, 6th Edition. (n.d.). Retrieved December 11, 2014, <http://www.mhs.com/product.aspx?gr=cli&prod=bvmi5&id=overview>
- Clinical Psychology. (n.d.). Retrieved December 11, 2014, from <http://www.pearsonclinical.com/psychology/products/100000663/the-beery-buktenica-developmental-test-of-visual-motor-integration-6th-edition-beeryvmi.html>
- Idoni, J., Taub, M. B., & Harris, P. A. (2014). A comparison of two tests of visual-motor integration. *Optometry & Visual Performance*, 2(4), 170-174.
- Krab, L.C., Aarsen, F. K., de Geode-Bolder, A., Catsman-Berrevoets, C.E., Arts, W.F., Moll, H.A., Elgersma, Y. (2008). Impact of neurofibromatosis Type 1 on school performance. *Journal of Child Neurology*, 23, 1002-1010.
- McCrimmon, A., Altomare, A., Matchless, R., Jitlina, K. (2012). Test review: The Beery developmental test of visual-motor integration. *Journal Of Psychoeducational Assessment*, 30(6), 588-592.
- Polumbinski, J., et al. (1986). Factor structure evidence for developmental levels of perceptual processing on the Developmental Test of Visual-Motor Integration. *Psychology in the Schools*, 23, 337-341.
- Snow, C.E., et al. (1998). *Preventing reading difficulties in young children*. Washington DC: National Academy Press.