

rated needs to be verified with data from samples sufficiently large to yield stable estimates. In future revisions, the test authors may wish to review the reading level and degree of understanding of formal linguistic conventions required of raters to ensure that the language is accessible to a wider range of raters, particularly those who are not extremely well-educated. Terms like “idiosyncratic” that appear in item examples and “past participle” and “indefinite adjective” found in item-level behavior statements seem more sophisticated than one could reasonably expect of a parent with a typical level of education who is living in the United States. Until then, as suggested in the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014), users are encouraged to exercise caution when interpreting scores based on ratings of children/adolescents by parents and teachers who differ demographically from the populations sampled in the studies described in the technical manuals and related research articles.

**SUMMARY.** The PDDBI provides reliable and valid scores that serve as measures of both adaptive and maladaptive behaviors commonly seen in individuals with ASD. Age-normed on children and adolescents diagnosed with ASD, the PDDBI is perhaps uniquely useful in tracking changes in behavior over time among people with ASD. The addition of the Autism Spectrum Disorder Decision Tree, to be used in combination with the PDDBI extended version (parent or teacher rating form), expands the utility of the PDDBI to the diagnostic arena. The ASD—DT shows good sensitivity and specificity, particularly when based on the ratings of parent informants for school-aged children. As with any norm-referenced assessment, users are urged to carefully consider the characteristics of the norming and standardization samples to determine whether they adequately represent the population about whom the user intends to make decisions.

#### REVIEWER'S REFERENCES

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: AERA.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5<sup>th</sup> ed.). Washington, DC: Author.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4<sup>th</sup> ed.). Washington, DC: Author.
- Baio, J., Wiggins, L., Christensen, D. L., Maenner, M. J., Daniels, J., Warren, Z., ... Dowling, N. F. (2018, April). Prevalence of autism spectrum disorder among children aged 8 years: Autism and developmental disabilities monitoring network, 11 sites, United States, 2014. *Surveillance Summaries*, 67 (6), 1-23. Retrieved from <http://dx.doi.org/10.15585/mmwr.ss6706a1>
- Burkett, K., Morris, E., Manning-Courtney, P., Anthony, J., & Shambley-Ebron, D. (2015). African American families on autism diagnosis and treatment: The influence of culture. *Journal of Autism and Developmental Disorders*, 45, 3244-3254.

Durkin, M. S., Maenner, M. J., Meaney, F. J., Levy, S. E., DiGuseppi, C., Nicholas, J. S., ... Schieve, L. A. (2010). Socioeconomic inequality in the prevalence of autism spectrum disorder: Evidence from a U.S. cross-sectional study. *PLoS ONE*, 5(7). doi:10.1371/journal.pone.0011551

Humes, K. R., Jones, N. A., & Ramirez, R. R. (2011, March). *Overview of race and Hispanic origin: 2010*. Retrieved from <https://www.census.gov/prod/cen2010/briefs/c2010br-02.pdf>

Nowell, K. P., Brewton, C. M., Allain, E., & Mire, S. S. (2015). The influence of demographic factors on the identification of autism spectrum disorder: A review and call for research. *Review Journal of Autism and Developmental Disorders*, 2, 300-309.

[ 117 ]

### Peabody Picture Vocabulary Test, Fifth Edition.

**Purpose:** Designed to assess receptive vocabulary for children and adults.

**Population:** Ages 2-6 to 90+.

**Publication Dates:** 1959-2019.

**Acronym:** PPVT-5.

**Scores:** Total score only.

**Administration:** Individual.

**Forms, 2:** A, B.

**Price Data, 2020:** \$405 per complete kit (A & B) including manual (2019, 343 pages), 2 stimulus books (A & B), and 25 of each record form (A & B); \$225 per individual kit (A or B); \$99 per manual; \$99 per stimulus book (A or B); \$49 per 25 record forms (A or B); \$40 per 1-year scoring subscription; \$4.50 per digital administration (A or B); \$1.10 per digital score summary report (quantity discounts available).

**Time:** Approximately 10-15 minutes for administration.

**Comments:** Co-normed with the Expressive Vocabulary Test, Third Edition (see 58, this volume) to allow comparison of a test taker's receptive and expressive vocabulary skills; available in print and digital formats.

**Author:** Douglas M. Dunn.

**Publisher:** Pearson.

**Cross References:** For reviews by Joseph C. Kush and Steven R. Shaw of the fourth edition, see 18:88; for reviews by Frederick Bessai and Orest Eugene Wasyliv of the third edition, see 14:280; see also T5:1903 (585 references) and T4:1945 (426 references); for reviews by R. Steve McCallum and Elisabeth H. Wiig of the revised edition, see 9:926 (117 references); see also T3:1771 (301 references), 8:222 (213 references), T2:516 (77 references), and 7:417 (201 references); for reviews by Howard B. Lyman and Ellen V. Piers of the original edition, see 6:530 (21 references).

*Review of the Peabody Picture Vocabulary Test, Fifth Edition by GARY L. CANIVEZ, Professor of Psychology, Department of Psychology, Eastern Illinois University, Charleston, IL:*

**DESCRIPTION.** The Peabody Picture Vocabulary Test, Fifth Edition (PPVT-5) is the fifth version of a popular measure of receptive vocabulary for children and adults (ages 2 years 6 months to 90+ years) that originated in 1959 (Dunn, 1959). It is a

major revision of the Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4; Dunn & Dunn, 2007). The PPVT-5 measures English vocabulary word knowledge (nouns, verbs, attributes) and was co-normed with the Expressive Vocabulary Test, Third Edition (EVT-3; Williams, 2019; see 58, this volume) for broad assessment and comparisons of receptive (listening) versus expressive (spoken) vocabulary skills in clinical and school applications (screening, diagnosis, research). It includes a detailed manual, two stimulus books (Forms A and B), and two record forms (A and B). In addition to the traditional administration and scoring, the test publisher provides digital administration, scoring, and reporting via two tablets on the Q-interactive platform and computerized scoring and reporting of traditional administration on the Q-global platform.

As an individually administered measure of receptive vocabulary, the PPVT-5 requires the examinee to choose one of four pictures corresponding to the word verbally presented by the examiner. It is untimed (although examiners are instructed to prompt a response after approximately 10 seconds), and median completion time ranged from 11 to 16 minutes during standardization. Administration is relatively simple, and scoring is objective. The test manual specifies guidelines for proper administration and scoring by professionals or paraprofessionals, but interpretation is reserved for properly trained professionals. The test manual provides useful descriptions of various testing environments, cultural diversity, and testing individuals with special needs to maximize valid assessment.

Detailed instructions for administration are provided in the test manual and include training items to teach the examinee how to respond and determine whether they may be properly assessed, starting points based on age, the basal rule (and reverse order testing), and the ceiling (discontinue) rule. Item administration is clearly presented instructing the examiner to say, "Put your finger on [word]," "Show me [word]," "Point to [word]," "Where is [word]?" and so forth. Additional instructions are provided for encouraging responses, guessing, and handling unclear responses. Additional accommodations are provided to facilitate assessment of very young children or those with physical disabilities that would prevent them from pointing. The two stimulus books are well organized with tabs marking training pages and item start points by age for efficient presentation and include high-quality

images. The two record forms are well designed for easy completion and include color coding of correct responses and pronunciation guides for the most difficult words.

Scoring the PPVT-5 is well described and illustrated with figures as is the process of transforming raw scores to standard scores ( $M = 100$ ,  $SD = 15$ ) with 90% and 95% confidence intervals, percentile ranks, normal curve equivalents, stanines, age equivalents (four major limitations explicitly described in the test manual), and growth scale values. Graphical presentation and qualitative descriptions are also provided and illustrated. Statistical significance of growth scale values across time can be determined as can differences between standard scores on the PPVT-5 and the EVT-3. When significant differences are observed between the PPVT-5 and the EVT-3, population base rates are provided for judging rarity.

**DEVELOPMENT.** The development of the PPVT-5 began with a survey of PPVT-4 users regarding test administration instructions, item words, and pictures. International test developers in various countries reviewed items for cultural differences. This feedback provided guidance for PPVT-5 item development. Newly created items ( $N = 109$ ) were examined for content and bias (gender, culture/ethnic, region, socioeconomic) through review by a panel of experts, and items identified by more than one panel member as potentially biased were eliminated. Pilot research was conducted on new items with a sample of 217 typically developing children using classical test theory (CTT), item response theory (IRT), item logistic regression, and differential item functioning (DIF) to assess adequacy for inclusion. All item stimulus art was replaced with newly created images designed for digital presentation. The test manual lists the number of items within 20 different content categories (actions, adjectives, animals, body parts, buildings, household objects, etc.) within Forms A and B, each containing 251 items for standardization.

#### TECHNICAL.

**Standardization.** Standardization of the PPVT-5 included digital administration by 217 examiners in 44 states who met specified criteria and were approved following submission of a practice test for quality control. The standardization sample was demographically representative of the U.S. population as per U.S. Census estimates in 2017, stratified by key variables of age, sex, race/ethnicity, geographic region, and parent education

level (proxy for SES); the sample included English-speaking individuals ages 2 years 6 months to 90+ years. A table in the test manual presents normative sample percentages for demographic variables of education, race/ethnicity, and region by age group and U.S. population percentages for comparison. There are no tables presenting percentages of education by race/ethnicity, education by region, or race/ethnicity by region, which would be useful. The normative sample included 2,720 participants meeting inclusion criteria for standard administration without modifications. It was noted that during standardization, participants first completed the PPVT-5 Form A followed by the EVT-3 Form A *or* the PPVT-5 Form B followed by the EVT-3 Form B, but it is unclear whether all 2,720 participants were administered *both* Forms A and B of the PPVT-5 and EVT-3, and if so, whether any form of counterbalancing was used and what the inter-form test interval was. If the standardization sample was randomly bifurcated such that 1,360 participants were administered PPVT-5 and EVT-3 Forms A and 1,360 participants were administered PPVT-5 and EVT-3 Forms B, it would reflect a sizable reduction in normative samples of *each* form and should be explicitly stated.

Each PPVT-5 form was reduced to 240 items for norm development using statistical analyses and feedback from examiners. As with other tests published by Pearson, such as the Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V; Wechsler, 2014), “inferential norming” (manual, p. 33) was used in which distributions of means, standard deviations, and skewness for each age group were examined from first- through fourth-order polynomial regressions with comparison to theoretical distributions and growth curves that produced percentiles for conversion to standard scores ( $M = 100$ ,  $SD = 15$ ). Distribution irregularities were reportedly corrected through smoothing, but the smoothing method (statistical vs. hand/visual) was not noted. One-parameter IRT (Rasch) models were used to produce ability scores that were rescaled into growth scale values ( $M = 500$ ,  $SD = 25$ ) for score comparisons across time. Critical values for statistical significance were produced using  $p = .10$  without explicit justification.

**Reliability.** Three types of reliability estimates for PPVT-5 scores are reported: internal consistency, alternate form, and test-retest stability. Strong evidence was provided in each case. Internal con-

sistency estimates by age group were produced by Spearman-Brown corrected split-half correlations ranging from .94 to .98 ( $M = .97$ ) and used for producing standard errors of measurement and confidence intervals. These should be considered best-case estimates because they do not consider other major sources of error such as long-term temporal stability, administration errors, or scoring errors (Hanna, Bradley, & Holen, 1981) known to influence test scores in clinical assessments, although given the ease of administration and scoring of the PPVT-5, such sources of error are likely less problematic than intelligence tests. Estimated true score confidence intervals (90% and 95%) are provided in the test manual and present increasingly asymmetrical confidence intervals the further from the mean the score is due to regression to mean effects. However, obtained score confidence intervals are appropriate (Glutting, McDermott, & Stanley, 1987; Sattler, 2008) when the assessment question is concerned with estimating the true score of the individual at the time of the evaluation (rather than the long-term estimate).

Alternate forms reliability (equivalence) was assessed with 273 participants in three age groups using counterbalancing of forms with random assignment of the first form administered and test intervals ranging from 0 to 34 days (mean and/or median not disclosed). Equivalence estimates were good ( $M = .86$ ): .81 for ages 2:6-11:11 ( $n = 132$ ), .86 for ages 12:0-24:11 ( $n = 75$ ), and .89 for ages 25:0-90:11 and older ( $n = 66$ ). Effect sizes (Cohen's  $d$ ) for mean differences between the two forms were trivial (.06-.07).

Short-term test-retest stability was investigated for 213 participants in three age groups with a mean retest interval of 30 days (range not disclosed). Stability coefficients ( $M = .84$ ) were good: .86 for ages 2:6-11:11 ( $n = 96$ ), .75 for ages 12:0-24:11 ( $n = 62$ ), and .89 for ages 25:0-90:11 and older ( $n = 55$ ). Effect sizes for mean differences between the two administrations were trivial to small (Cohen's  $d = .08$ -.32,  $M = .16$ ).

**Validity.** Support for validity of PPVT-5 scores was focused on evidence based on test content, response processes, relationships with other variables, and special group studies (distinct group differences). Review of literature, test user feedback, and expert reviews were reported as sources of evidence for test content. Empirical and qualitative methods were noted for assessing evidence based

on response processes in pilot testing of new items as well as standardization items. Evidence based on relationships with other variables (convergent and discriminant validity) included comparisons of PPVT-5 scores with scores from the PPVT-4 ( $M = .79$ ), with language based tests such as the EVT-3 ( $M = .76$ ), CELF Preschool-2 ( $r$ s ranged from .61 to .75), and CELF-5 ( $r$ s ranged from .68 to .73), and with the Kaufman Test of Educational Achievement-Third Edition Brief (KTEA-3 Brief) ( $r$ s ranged from .42 to .46). These results illustrate expected convergent and discriminant relationships.

Special group studies (distinct group differences) examined differences in PPVT-5 performance of individuals identified with language delay ( $N = 120$ ), specific language impairment ( $N = 100$ ), specific learning disability in reading and/or writing ( $N = 162$ ), hearing impairment ( $N = 70$ ), or autism spectrum disorder ( $N = 118$ ) with matched (age, sex, race/ethnicity, parent educational level) samples of typically developing individuals. Mean differences were large with typically developing youth showing average PPVT-5 scores and special groups scoring significantly lower with large effect sizes (Cohen's  $d$ ): language delay ( $d = 1.93$ ), specific language impairment ( $d = 1.68$ ), specific learning disability ( $d = .90$ ), hearing impairment ( $d = .89$ ), and autism spectrum disorder ( $d = .81$ ).

**Diagnostic Utility (Accuracy).** The test author keenly recognized that special group studies or group differences are necessary but not sufficient indicators of clinical utility and provided extensive description of diagnostic accuracy methods and the variety of important indicators for judging such accuracy described by Kessell and Zimmerman (1993). Using the language delay ( $N = 120$ ), specific language impairment ( $N = 100$ ), and specific learning disability in reading and/or writing ( $N = 162$ ) special groups and their matched samples, the test author reported the diagnostic accuracy of PPVT-5 scores across three cut scores (85, 77, and 70, corresponding to 1, 1.5, and 2  $SD$ s below the mean) as well as five different base rates, which are known to affect diagnostic utility statistics. Discussion of positive results focused primarily on sensitivity and specificity estimates, but in the context of correctly identifying those who truly have or do not have a disorder or condition, it is the positive and negative predictive power estimates presented that are more important.

**COMMENTARY.** The PPVT-5 is well designed and normed, and incorporation of digital

administration via Q-interactive may be a useful adaptation. However, the test manual did not include evidence for the equivalence of scores produced from the digital versus the traditional method of administration, and such support is essential. Also, while the inclusion of diagnostic accuracy was welcome and refreshing and it was acknowledged that base rates and cut scores affect such estimates, one method that should be considered in future studies and revision is the use of receiver operating characteristic (ROC) curve analysis (Swets, 1996; Treat & Viken, 2012), which is not affected by base rates or cut scores. As a very narrow measure of receptive vocabulary the test's value will mostly be for screening or supplementing assessments; however, vocabulary knowledge is an important component of speech and language and reading, and a narrow aspect of general intelligence. Speech and language pathologists, school psychologists, early childhood specialists, reading specialists, and clinical psychologists will find the PPVT-5 familiar and useful given preliminary supportive evidence for reliability, validity, and diagnostic utility. Such positive results should be replicated and extend these preliminary results.

**SUMMARY.** The present revision represents a well-designed, relatively inexpensive, easy-to-use measure of receptive vocabulary that has good norms and ample preliminary evidence for score reliability, validity, and utility. As such the PPVT-5 is an excellent revision that should provide users confident assessment of receptive vocabulary for comparison to other skills and attributes in screening, clinical assessment, and research.

#### REVIEWER'S REFERENCES

- Dunn, L. M. (1959). Peabody Picture Vocabulary Test. Circle Pines, MN: American Guidance Service.
- Dunn, L. M., & Dunn, D. M. (2007). Peabody Picture Vocabulary Test, Fourth Edition. San Antonio, TX: Pearson.
- Glutting, J. J., McDermott, P. A., & Stanley, J. C. (1987). Resolving differences among methods of establishing confidence limits for test scores. *Educational and Psychological Measurement*, 47, 607-614.
- Hanna, G. S., Bradley, F. O., & Holen, M. C. (1981). Estimating major sources of measurement error in individual intelligences scales: Taking our heads out of the sand. *Journal of School Psychology*, 19, 370-376.
- Kessell, J. B., & Zimmerman, M. (1993). Reporting errors in studies of the diagnostic performance of self-administered questionnaires: Extent of the problem, recommendations for standardized presentation of results, and implications for the peer review process. *Psychological Assessment*, 5, 395-399. doi:10.1037/1040-3590.5.4.395
- Sattler, J. M. (2008). *Assessment of children: Cognitive foundations* (5th ed.). San Diego, CA: Author.
- Swets, J. A. (1996). *Signal detection theory and ROC analysis in psychological diagnostics: Collected papers*. Mahwah, NJ: Erlbaum.
- Treat, T. A., & Viken, R. J. (2012). Measuring test performance with signal detection theory techniques. In H. Cooper, P. M. Camie, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Volume 1: Foundations, planning, measures, and psychometrics* (pp. 723-744). Washington, DC: American Psychological Association.
- Wechsler, D. (2014). *Wechsler Intelligence Scale for Children-Fifth Edition technical and interpretive manual*. San Antonio, TX: Pearson.
- Williams, K. T. (2019). *Expressive Vocabulary Test, Third Edition*. San Antonio, TX: Pearson.

*Review of the Peabody Picture Vocabulary Test, Fifth Edition by THERESA GRAHAM LAUGHLIN, Adjunct Faculty, Nebraska Methodist College of Nursing, Omaha, NE:*

**DESCRIPTION.** The Peabody Picture Vocabulary Test, Fifth Edition (PPVT-5) is a norm-referenced assessment tool designed to measure receptive vocabulary of Standard American English for individuals ages 2 years 6 months to 90 years and older. A feature of the PPVT-5 is that it was co-normed with the Expressive Vocabulary Test, Third Edition (EVT-3; Williams, 2019). The PPVT-5 includes two parallel forms, A and B, consisting of 240 items on each form. The current edition was updated to include items that are shared between PPVT-5 Forms A and B. In addition, some items are shared between the PPVT-5 and the EVT-3, which contributes to the comparability of the two instruments. All of the artwork has been newly drawn to be “familiar to individuals of various ages and from diverse regional and cultural backgrounds” (manual, p. 1). Finally, the norming tables have been updated to represent demographic changes in the United States. Using pictures and verbal prompts, the PPVT-5 can be used as a screening tool to identify receptive language disorders in children and adults, to examine preschoolers’ vocabulary development, to provide a tool for reevaluation, and to understand reading difficulties. The PPVT-5 is also available on a digital platform for both administration and scoring. Intervention reporting is available on the digital platform but is not included in this review.

The PPVT-5 materials include an examiner’s manual, stimulus books with pictures for Forms A and B, and examiner record forms for Forms A and B. The examiner’s manual provides information on administration, scoring, and interpreting the test. In addition, it describes the development and standardization of the current version of the PPVT-5 and includes evidence of reliability and validity. The appendices provide information on age equivalents and testing examinees with hearing loss. Finally, information is provided on item analyses on subsets of items in the areas of home versus school vocabulary, parts of speech, and STEM vocabulary.

The PPVT-5 is administered individually. The instrument is untimed, and starting points based on chronological age allow for quick administration (10–15 minutes). Examiners should be familiar with the testing materials and have experience working with individuals in the age group being tested, but

extensive examiner training is not necessary. If the PPVT-5 is administered along with the EVT-3, then the PPVT-5 should be administered first. The stimulus book is an easel-style book with each page displaying four pictures in a quadrant. Each picture is numbered (1–4) at the bottom. The record form includes a cover page with identifying information (e.g., name, age), raw scores and summary scores, and a graphical profile. Training and test items are listed with the target word on the record form. Examinees are instructed that they will hear a target word and that they can either point to the picture of the target word or indicate the number corresponding to the picture of the target word. Scoring is indicated on the record form by circling the number that the examinee indicates either from gesture or verbal response. The correct answer is indicated on the record form in red. The examiner can then indicate the accuracy of the answer at the time of administration. If the examinee does not respond, then the item is coded as incorrect and “NR” is circled, indicating there was no verbal or gestural response. A basal is established by three consecutive correct responses. A ceiling is established by six consecutive incorrect responses. Because responses are scored as the examinee answers, testing ends once the ceiling has been reached.

The raw score is based on the total number of errors subtracted from the total number of items attempted. Tables are provided in the appendix of the test manual to convert raw scores to standard scores. In addition, the appendix provides the confidence intervals, percentile ranks, normal curve equivalents, and stanines, all of which are transferred to the front page of the record form. The graphical profile provides an additional view of the examinee’s performance. The last page of the record form provides space to compare PPVT-5 scores over time and to compare performance on the PPVT-5 with performance on the EVT-3. Tables are provided in the appendix of the test manual to evaluate whether changes in scores on the PPVT-5 or comparisons with EVT-3 scores are statistically significant.

**DEVELOPMENT.** New items were developed through a review of high-frequency words for young children and words that may be used in home, community, and school settings for older examinees. An item was considered new if the word had not been included in the PPVT-4 or if 50% or more of the images on its stimulus page were different or had been significantly modified

from the PPVT-4. A total of 109 new items were embedded with 93 items from the PPVT-4 for the pilot study, which included 217 individuals with no known language disorder or hearing deficit ranging in age from 2 years 6 months to 21 years 11 months. The reported results from the pilot research were a bit vague indicating that “the majority of new items were familiar to the individuals who were presented them” (manual, p. 30). A total of 16 new items were added, and 26 items from the PPVT-4 were omitted. No specific information is provided to explain why items were omitted nor to which form the new items were added. In addition, no information was provided on whether the distribution of new items or omission of previous items was equal across age groups.

The examiner’s manual provides information on interpreting quantitative scores (e.g., standard scores, growth scale values) from the PPVT-5. In addition, information on qualitative analyses was provided. The PPVT-5 was designed for further item analysis in the areas of home versus school vocabulary, vocabulary by part of speech, and STEM vocabulary. However, little information was provided as to how those distinctions were determined, and some words designated as “home” or “school” easily could be argued to belong in the other category. Without an explanation, it is unclear how helpful these additional analyses can be. Although information is provided regarding interpretation of scores, it is clear that although administration can be accomplished by an experienced but not expert test administrator, interpretation of results requires a much more thorough analysis by an expert in the field of language development.

**TECHNICAL.** The standardization study included 2,720 individuals ranging in age from 2 years 6 months to 90 years and older recruited from 44 states. Most age bands consisted of 100 individuals stratified by age, sex, race/ethnicity, geographic region, and parent/caregiver education level within 5% of Census distributions for race/ethnicity and parent/caregiver education level. Examiners included speech-language pathologists, psychologists, and individuals experienced with administering standardized tests. A total of 217 examiners completed and submitted practice test protocols that were evaluated prior to their participation in the standardization study. Start points were based on whether 90% or more of the examinees in a particular age group passed an item. Rasch ability scores were used to develop rules to establish

basals and ceilings. Standard scores with a mean of 100 and a standard deviation of 15 were developed using inferential norming (Zhu & Chen, 2011).

**Reliability.** Reliability of the PPVT-5 was evaluated using three measures: internal consistency, alternate form reliability, and test-retest stability. Internal consistency was examined using the split-half method for both Forms A and B within each age band and resulted in very high coefficients (.94-.98). Internal consistency also was evaluated by providing standard error of measurement confidence intervals. SEMs ranged from 2.12 to 3.67, further indicating a high degree of reliability of the PPVT-5 scores.

Alternate form reliability was assessed by comparing the performance of 273 examinees who took both Forms A and B. Results were compared across three age bands. Although it is not clear how these age bands were determined, reliabilities between the two forms were high, with correlation coefficients ranging from .81 to .89.

Test-retest stability was examined in a study using only Form A in which 213 individuals were administered the PPVT-5 twice with approximately a one-month retest interval. Although the second administration resulted in higher average scores (as expected from practice), the correlation coefficients (corrected for variability) were high, ranging from .86 to .89.

**Validity.** Evidence of validity for the PPVT-5 score was established by reviewing the test content, its relationship to other variables, special group studies, and diagnostic accuracy. Test content was evaluated through literature review, users’ feedback, and expert review. Feedback from test users and test developers from the test publisher’s international offices was obtained. Interestingly, feedback was solicited from many international contacts but did not include a representative from Mexico. Given that the standardization sample comprised greater than 20% of individuals identifying as Hispanic, and that one of the stated goals of the current revisions was to create items that are familiar to a broader cultural background, it is unfortunate that a broader Hispanic perspective was not included. [Editor’s note: The test publisher advised the Buros Center for Testing that a panel of 12 consultants, including four with expertise in Hispanic culture, reviewed the items to ensure their appropriateness regarding gender, ethnicity, regional differences, and varying socioeconomic levels.]

To examine the relationship of the PPVT-5 to other measures, including the EVT-3, PPVT-4,

Clinical Evaluation of Language Fundamentals–Fifth Edition (CELF-5), CELF Preschool-2, and Kaufman Test of Educational Achievement–Third Edition Brief (KTEA-3 Brief), smaller sub-studies were conducted. Other than the comparison with the EVT-3 (which was co-normed with the PPVT-5), the sample sizes of the smaller studies ranged from 56 participants (CELF Preschool-2) to 208 participants (KTEA-3 Brief). High correlation coefficients were found between the PPVT-5 and both the PPVT-4 (range = .76-.81) and the EVT-3 (range = .75-.77). The standard difference ranged from .10 to .33 with the PPVT-4 and was .01 with the EVT-3. The test authors suggest the higher outcome with the PPVT-4 may be due to differences in the normative population creating a broader range of ability in younger children on the PPVT-5. Moderate correlation coefficients were found between the PPVT-5 and the CELF-5 (.68-.73) and the CELF Preschool-2 (.61-.75). The PPVT-5 was also moderately correlated with the KTEA-3 Brief but with lower correlation coefficients (.42-.46) than were noted with some of the other measures.

Although the PPVT-5 was not designed to be a diagnostic test for different language delay disorders, performance on the PPVT-5 should be able to differentiate individuals with typical vocabulary development from different clinical subgroups whose receptive vocabulary may be delayed. Multiple studies were conducted to ascertain whether PPVT-5 scores would differentiate typically developing individuals from those with language impairment, learning disability, autism spectrum disorder, and hearing impairment. In each study, the test author used a matched control and found that the PPVT-5 correctly discriminated between typically developing individuals and those with language delay, specific language impairment, specific learning disability in reading/writing, autism spectrum disorder, and hearing impairment with cochlear implants.

The test author further examined the data from the sub-studies on language delay, specific language impairment, and learning disability in reading/writing to explore whether scores on the PPVT-5 could identify individuals with a language disorder relying on both positive and negative predictive power models. Results suggested that the PPVT-5 has high sensitivity and specificity and therefore can be used to identify individuals with language delay when used with other measures.

**COMMENTARY AND SUMMARY.** PPVT-5 is a norm-referenced assessment of

receptive vocabulary of individuals ages 2 years 6 months to 90 years and older that can be completed in about 15 minutes. Because the PPVT-5 is conormed with the EVT-3, administration of the two tests provides a snapshot of both receptive and expressive vocabulary. The standardization sample sufficiently represents current U.S. demographic trends. Although there are some concerns with the sample used in the sub-studies in terms of size and age distribution, the standardization study described in the examiner's manual provides sufficient evidence of reliability and validity. The PPVT-5 is not meant to be a diagnostic tool. Like its predecessors, it is best used as an initial screening for vocabulary development or used in conjunction with other measures of language ability.

#### REVIEWER'S REFERENCE

Zhu, J., & Chen, H. (2011). Utility of inferential norming with smaller sample sizes. *Journal of Psychoeducational Assessment*, 29(6), 570-580.

[ 118 ]

### Pearman Personality Integrator.

**Purpose:** Designed as "a measure of personality that gauges one's natural state (i.e., what is most comfortable) and one's everyday environment (i.e., what is most often demonstrated)."

**Population:** Ages 18 and older.

**Publication Date:** 2015.

**Administration:** Individual or group.

**Restricted Distribution:** Certification or pre-qualification required.

**Parts, 2:** Pearman Personality; Pearman FlexIndex.

**Price Data:** Available from publisher.

**Time:** 20-30 minutes for completion.

**Comments:** Administered and scored online; user's handbook only available online.

**Author:** Roger R. Pearman.

**Publisher:** Multi-Health Systems, Inc.

#### a) PEARMAN PERSONALITY.

**Scores, 42:** 14 Circle Scores: Natural Overall Attitude (Extraversion or Introversion), Demonstrated Overall Attitude (Extraversion or Introversion), Natural Overall Perceiving (Sensing or Intuiting), Demonstrated Overall Perceiving (Sensing or Intuiting), Natural Overall Judging (Thinking or Feeling), Demonstrated Overall Judging (Thinking or Feeling), Natural Extraverted Perceiving (Extraverted Intuiting or Extraverted Sensing), Demonstrated Extraverted Perceiving (Extraverted Intuiting or Extraverted Sensing), Natural Introverted Perceiving (Introverted Intuiting or Introverted Sensing), Demonstrated Introverted Perceiving (Introverted Intuiting or Introverted Sensing), Natural Extraverted Judging (Extraverted Feeling or Extraverted Thinking), Demonstrated Extraverted Judging (Extraverted Feeling or Extraverted Thinking), Natural Introverted Judging (Introverted Feeling or Introverted Thinking), Demonstrated Introverted