Section 5 –
Testing and Assessment

The regulation requires a score of ninth stanine for identifications in general intellectual ability and specific academic aptitude; however, KDE has not endorsed or required that any specific test(s) be used.

Section Includes:

- Role of Assessments in Identification – NAGC Position Paper
- Clarification 2007
- Commonly Used Testing Instruments in Gifted Education
- Using WISC-IV For Gifted Identification – a NAGC Position Paper
- Norm- and Criterion-Referenced Testing
- Connecting Performance Assessment to Instruction
- Creating Meaningful Performance Assessments

THE ROLE OF ASSESSMENTS IN THE IDENTIFICATION OF GIFTED STUDENTS

Assessments can be used for a variety of purposes, including identifying students for gifted programs; providing ongoing feedback to guide the instructional process; and to determine to what extent students have obtained intended goals (e.g., academic, affective) within a gifted program. The purpose of this position paper is to provide parents, teachers, and other advocates of gifted students with best practices endorsed by NAGC related to the first purpose--the role of assessments in identifying students for gifted programs.

NAGC believes that the process of identifying students for gifted and talented programs must be based on defensible measurement practices, including the process of selecting psychometrically sound assessments aligned with a program's goals and objectives; the administration and interpretation of the assessments by individuals with appropriate credentials or training; and the ethical application of decisions regarding gifted program placement. Further, NAGC believes that there are specific practices that are supportive of these measurement practices.

In recent years, there have been significant discussions regarding the role of traditional assessments in identifying students who are typically under-represented in gifted programs, including culturally and linguistically diverse and low-income gifted students, and the use of alternative assessments with these students such as nonverbal ability tests (Lohman, 2005). NAGC believes that assessments selected for use in the identification of gifted students must be sensitive to and appropriate for the characteristics of the students being assessed and must aim to be inclusive of students from different cultures, races, and economic circumstances. Program administrators should choose the most psychometrically sound assessments with appropriate norms for their population of students and programs and use them appropriately for selection (see Lohman, 2005). However, it is also imperative that test users and policymakers understand that alternative-type assessments are not panaceas to the issue of underrepresentation each come with limitations in terms of reliability and validity, and that these types of assessments should never be used in isolation to identify gifted children.

Another issue that warrants consideration in the identification of gifted students is the decision to use group versus individual testing, which is often determined by the availability of resources and the characteristics of the children to be evaluated. More accurate assessment data may be obtained via one-on-one testing with very young children and children with special characteristics and needs such as those with dual exceptionalities. For these children it is important to have a tester who is sensitive to and experienced with the group being assessed as well as the training in the administration of the assessments.

NAGC believes that because the use of assessments is an integral part of the identification process, test users have a responsibility to ensure that all testing is conducted in a fair and ethical manner. Such practices include the appropriate storing of testing materials before, during, and after testing; training all personnel involved with the administration and/or scoring of assessments; utilizing assessments that are developmentally appropriate and for only the purposes for which they were developed; interpreting assessment results to the appropriate audiences; and maintaining the confidentiality of students at all times. While NAGC advocates for the use of multiple assessments in the identification of gifted students, NAGC also believes that combining disparate data from multiple assessments must be done in such a way as to identify not only those students who are in immediate need of instruction beyond the regular curriculum, but also those students who display the potential for high-level learning beyond the regular curriculum.

In order to best implement defensible assessment practices for the purposes of gifted program identification, NAGC supports the collaboration of multiple stakeholders, including teachers, parents, and other advocates of gifted children, as well as general education administration at the district and state levels. This collaboration works to

ensure that the application of defensible measurement practices results in the equitable and consistent use of assessments for the purposes of gifted program identification.

Research-Based Practices Regarding the Use of Assessments for Identification Purposes

Regardless of the type of assessments used for identification or whether students are assessed in groups or individually, there are five non-negotiable practices in the use of assessments as identification tools. First, the choice of assessment tools must match the definition of giftedness that has been determined by the state, district, or school. The degree to which the assessment tool is aligned with the definition of giftedness is an important aspect of validity. Further, any assessments used in the identification process also should be aligned with the gifted program's goals and objectives and desired outcomes for students as a result of participation in the program (Feldhusen, Asher, & Hoover, 1984). Program administrators must carefully consider the program's goals and objectives as well as the aptitudes, achievement levels, and other characteristics of students (e.g., motivation, persistence, interest) needed for success in the program in order to select instruments that provide the most reliable and valid data regarding students' potential for success.

Second, identification of gifted and talented students should not be based on a single assessment. Rather, multiple pieces of evidence should be collected that measure different constructs and characteristics aligned to the gifted program's definition, goals, and objectives (Callahan, Tomlinson, & Pizzat, 1993), ideally including a variety of format types (e.g., paper-and-pencil; performance assessment). Multiple pieces of psychometrically sound data obtained from a variety of sources result in a more comprehensive and thus, more accurate picture of the student on which to base selection. For example, if trying to measure mathematical ability, appropriate choices might include a selected-response, domain-specific mathematics achievement test (e.g., a multiple-choice assessment) and a constructed-response assessment (e.g., performance assessment) where the student solves problems in an authentic context. However, when multiple assessments are used, it is important that the assessments provide different types of information as well as measure the construct, i.e. mathematical reasoning ability, differently. For example, although multiple pieces of information are being collected, administering assessments that follow the same response format may unfairly penalize some students while benefiting others. Program administrators should consider the use of a variety of format types when considering the specific assessments that will be used in an identification process and choose assessments sensitive to the inclusion of under-represented groups, culturally and linguistically diverse, and twice-exceptional students.

Third, the assessment conditions should mimic as closely as possible a natural setting in which the student can fully demonstrate his or her knowledge, skills, and abilities. The greater the unfamiliarity of the assessment setting, the greater the potential for undue negative influences on a student's performance (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 1999). For example, testing some of a district's second-grade students in a high school cafeteria on a given Saturday, while other second graders are administered the assessments within their classroom context, unfairly penalizes those students who are assessed outside their natural setting.

Fourth, school system personnel have the responsibility to be well-informed consumers regarding the technical documentation of each assessment used for identification (Joint Committee on Testing Practices, 2004). Assessment developers or publishers should include information on an instrument's psychometric properties (e.g., reliability and validity) and only assessments with adequate psychometric properties should be used in the identification of gifted students. In the absence of this information, responsible persons should determine an instrument's reliability and validity for diverse populations prior to using the instrument in an identification process.

Fifth, school system personnel have the responsibility to ensure that persons who administrator and score assessments used for identification are appropriately trained and that placement decisions are driven by defensible data and not based on personal relationships, political associations, or parental pressure.

The Variety of Assessment Types

Assessments differ on dimensions such as: the degree to which they are standardized (e.g., using large national samples versus local samples); the type of response format (e.g., producing a response as opposed to selecting a response from a predefined set); the ways in which the material is presented (e.g., paper-and-pencil, computerized, oral); and the content (e.g., mathematics) or constructs (e.g,. creativity) being assessed. NAGC believes that regardless of the type of assessment, only assessments that provide psychometrically sound information on students, regardless of language, culture, gender, race, or socio-economic status, should be used. The following are three types of assessments often used in identifying students for participation in programs and services for gifted learners.

1. ***Objective-type instruments:*** These types of selected-response assessments used for identification purposes range from standardized, nationally normed paper-and-pencil or computerized tests to locally developed and normed tests, including most of the aptitude and achievement tests used in schools as well as IQ tests (see NAGC position paper; "Use of WISC-IV for Gifted Identification"). When using these types of assessments, users should be fully aware of the test's purposes and have evidence of sufficient reliability of the test scores. In addition, test users should use assessments that have a sufficient ceiling for measuring students' aptitudes or achievement, lack item bias, and have support for the validity of the types of decisions that will be made based on the results of the assessment (Joint Committee, 2004).

2. ***Performance assessments****:* Performance assessments, authentic assessments, and portfolios are constructed-response assessments that may be used in the identification process. These types of assessments directly measure the domain-specific construct of interest. Examples of performance assessments include open-ended or extended-response items. For example, students might be asked to present arguments for or against a particular position on an issue, write in response to a prompt, or conduct and write a report of a scientific investigation. Portfolios are examples of another type of performance assessment in which students present their 'best pieces' highlighting the strengths of each piece or a 'work in progress' where students illustrate their improvement over time. When using these types of assessments, test users have the responsibility of ensuring that high-quality training procedures for scoring students' responses or rating students' work are in place in order to achieve a sufficient standard for exact rater agreement (Moon & Hughes, 2002). The acceptable standard for rater agreement is 80% exact agreement between two raters evaluating the same student response.

3. ***Rating Scales, Interviews:*** Classroom observations of students' behaviors, collected by the use of rating scales designed to assess student characteristics or behaviors, and student interviews can provide useful supplemental data, particularly on students whose talents may not be evident on traditional aptitude or achievement tests. NAGC believes that the use of rating scales and interviews should play only a supplementary role in the identification process. Collecting these types of information is very difficult to do well because all individuals are affected by bias and prejudice, even if only at a subconscious level. If these types of data are collected, it is important that one recognize that different genders, cultures, races, ethnicities, and social classes have different ways of communicating which may impact an observer's/interviewer's perspective on what behaviors constitute giftedness. It is also essential to recognize one's own views and predispositions relative to these differing subgroups of the population. To guard against the introduction of observer/interviewer bias into the identification process, educators should use structured tools with inclusive, but specific and clear, criteria to guide the data collection process (Oosterhof, 2003). Program administrators have the responsibility to ensure that individuals collecting these types of data have sufficient training in both the use of the instrument as well as the manifestation of giftedness in differing subgroups.

Implications for Practice

Program administrators are responsible for ensuring that:

1. the identification process and the assessments used are aligned with the program's definition of giftedness;
2. the process includes the use of multiple assessments that are combined in a reasoned way that is not biased against any particular subgroup of students (VanTassel-Baska, 2007);
3. the types of assessments used have sufficient psychometric evidence supportive of decisions about students' readiness for gifted programming;
4. all individuals involved in the assessment process have sufficient training in the administration and use of the assessments;
5. they themselves are fully informed about best practices in the field of testing as well as the latest research regarding the identification of gifted students; and
6. there is a process in place whereby the identification process is periodically evaluated to ensure it is reflective of best practices in the identification of gifted students.

Approved October 2008

Annotated Bibliography

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

This reference outlines the standards associated with the testing process. The reference provides the criteria for the evaluation of tests, testing practices, and the consequences of using tests and is applicable for those who develop tests; those who select or review tests; those who administer and score tests; and those who use the results from tests for decision-making purposes, among others.

Callahan, C. M., Tomlinson, C. A., & Pizzat, P. M. (Eds.). (1993). *Contexts for promise: Noteworthy practices and innovations in the identification of gifted students.* Charlottesville: National Research Center on the Gifted and Talented, University of Virginia.

This report is based on a National Research Center on the Gifted and Talented study that involved the collection and evaluation of instruments and processes used in identification processes from districts across the nation. The work includes many of the projects that have been funded by the Jacob K. Javits Gifted and Talented Students Education Act.

Feldhusen, J. F., Asher, J. W., & Hoover, S. M. (1984). Problems in the identification of giftedness, talent, or ability. In J. S. Renzulli (Ed.) (2004), *Identification of students for gifted and talented programs (*pp. 79-85), Thousand Oaks, CA: Corwin Press.

This article, first appearing in *Gifted Child Quarterly* in 1984, discusses five steps associated with the identification of students for gifted programs: defining of program goals and the types of students to be served; procedures for nomination; procedures for assessment; individual differentiation; and validation of the identification process.

Joint Committee on Testing Practices. (2004). *Code of fair testing practices in education.* Washington, DC: American Psychological Association.

The reference is a collaborative effort among the American Counseling Association, the American Educational Research Association, the American Psychological Association, the American Speech-Language-Hearing

Association, the National Association of School Psychologists, the National Association of Test Directors, and the National Council on Measurement in Education. The Code serves as a guide for individuals who are test developers as well as test users and focuses on the (1) development and selection of assessments; (2) administration and scoring of assessments; (3) reporting and interpreting assessment results; and (4) informing of test takers.

Lane, S., & Stone, C. A. (2006). Performance assessment. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed.), (pp. 387-432). National Council on Measurement in Education & American Council on Education. Westport, CT: Praeger Publishers.
This chapter provides an overview on the design, use, and validity of performance assessments for large-scale educational testing.

Lohman, D. F. (2005). The role of nonverbal ability tests in identifying academically gifted students: An aptitude perspective. *Gifted Child Quarterly, 49*, 111-138.
This article presents an overview of the different types of nonverbal ability tests, from individually administered to group administered and how selecting students on the basis of these types of assessments alone excludes many students who would profit from gifted and talented programs and includes many students for whom gifted and talented programs would be an ill fit.

Moon, T. R., & Hughes, K. R. (2002). Training and scoring issues involved in large-scale writing performance assessments. *Educational Measurement: Issues and Practice*, *21*(2), 15-19.
This study investigated the amount of error introduced into students' scores from constructed-response items based upon the type of training raters received as well as the type of scoring used for evaluating student responses.

Oosterhof, A. (2003). *Developing and using classroom assessments.* Upper Saddle River, NJ: Pearson Education.
This text provides a thorough and succinct discussion of the issues involved with using assessments in an educational environment, from development of to the uses of to the interpretation of various types of assessments as well as the issues that need to be taken into consideration when using each type of assessment.

VanTassel-Baska, J. (2007). *Alternative assessments with gifted and talented students.* Waco, TX: Prufrock Press.
This text provides an introductory guide to the methods used in educational settings for identifying gifted students as well as discussion for the need to identify students from underrepresented populations for gifted and talented programs. It also focuses on ways to assess the learning of gifted student through alternative means.

Clarification Regarding Testing for Gifted Identification

Leah Ellis
January 16, 2007

RE: Norm-referenced, standardized testing and using KCCT for identification of gifted students

This information is intended to provide guidelines for interpreting 704 Kentucky Administrative Regulation (KAR) 3:285. Programs for the gifted and talented. The Kentucky Department of Education is here to assist and ensure the implementation of this interpretation and the gifted regulation.

The regulation for the identification of gifted students in specific academic area(s) states:

704 KAR 3:285. Programs for the gifted and talented.
Section 3. Identification and Diagnosis of Gifted Characteristics, Behaviors, and Talent and Determination of Eligibility for Services

(b) Specific academic aptitude shall be determined by composite scores in the ninth stanine on one (1) or more subject test scores of an achievement test. If a student scores low on a formal group measure of academic strength, yet other documentation shows potential, the district shall administer another standardized normed achievement test. Evidence of specific academic aptitude also may include:

1. High performance on an additional individual or group test of academic aptitude;
2. Student awards or critiques of performances;
3. Off-level testing;
4. Portfolio of high academic performances; or
5. Student progress data.

The Kentucky Core Content Test (KCCT) is not a standardized, nationally normed-referenced test (NRT) as required in the regulation for the identification of specific academic aptitude giftedness. KCCT is a criterion-referenced test and should not be used for the formal test score to identify specific academic giftedness. However, the KCCT can be used as supporting evidence as "High performance on an additional individual or group test of academic aptitude."

Prior CATS (Commonwealth Accountability Testing System) results did include CTBS (Comprehensive Tests of Basic Skills), which is acceptable to use if a student obtains a composite score in the 9th stanine. Please note that CTBS subtests, such as vocabulary or math computation, are not composite scores.

Although KDE has offered guidance in the past on the acceptable use of the KCCT for identification of gifted students, our current focus is on the regulation and its intent. The Office of Teaching and Learning is collaborating with the Office of Assessment and Accountability to address this specific issue of making our assessment regulations and available testing tools more "gifted-friendly."

While there will not be a NRT for accountability purposes at the elementary or middle school level this year, 2006-2007, the Kentucky Board of Education (KBE) is considering having all elementary schools administer a NRT in reading and mathematics at some point in the elementary level. These discussions will continue at the February 2007 KBE meeting. Once the Board has made a final decision, KDE will notify districts and schools of the decision, including any decisions related to providing funding assistance to support districts toward the purchase of a NRT.

KDE encourages districts to look at their policies and procedures and rethink the process of identification from a different perspective. For example, when focusing on a target population, i.e. 3rd or 4th grade, a district may use a battery of screening tools that may include the KCCT, teacher checklists, referrals, student inventories, parent surveys, diagnostic classroom activities, peer surveys, portfolios, etc. The resulting smaller group of students, with the collected evidence from screening, would then be tested with a purchased NRT as a last piece of data after the other evidence is in place.

Use of the WISC-IV forGifted Identification

School districts use multi-faceted approaches to identify gifted students. Some states and districts employ comprehensive individual IQ tests as one of several identifiers. The most popular of these is the *Wechsler Intelligence Scale for Children, Fourth Edition* (WISC-IV) (Lubin, Wallis & Paine, 1971). Even in districts where IQ tests are not used in student selection, the WISC-IV is often administered when the parents appeal the decision to deny a child services.

Also, for twice exceptional children, the WISC-IV plays an important role in documenting the child's giftedness and learning deficits, as well as revealing the giftedness of children with expressive, physical, or other disabilities. In prior versions of the Wechsler scales, the child's Full Scale IQ score has been the primary determining factor in placement. However, the Full Scale IQ score of the WISC-IV often does not represent a child's intellectual liabilities as well as the General Ability Index. Therefore, some guidelines for test interpretation are necessary.

This position statement is designed for school psychologists, coordinators of gifted programs, teachers, and all professionals who determine placements based on IQ scores or design services based on a child's strengths and weaknesses. It is also provided for parents so they can better understand the interpretation of their children's scores. It is not intended to narrow the choice of tests in the selection of gifted students, but to broaden the guidelines for use of the WISC-IV and prevents its use in a way that is disadvantageous to gifted children.

The WISC-IV was standardized on 2200 children, including Caucasians, African Americans, Hispanics, Asians, and others (a combined designation including Native Americans, Alaskan Natives, and other groups in the U.S.), in proportion to their distribution in the American population. Parental educational levels and geographic regions were also proportionately represented. In concert with the publishers' concerns for "Suitability and Fairness," greater flexibility is built into the administration of the WISC-IV: examiners are permitted to use appropriate substitutions of subtests when necessary for equitability (Wechsler, 2003). Nevertheless, IQ tests should be interpreted cautiously for children from culturally and linguistically diverse backgrounds, and for all children, and should never be the only basis for exclusion from gifted programs. In addition, all efforts should be made to accommodate linguistic diversity and test children in their native language.

The WISC-IV introduces important structural changes that compromise the relevance of the Full Scale IQ score (FSIQ) for gifted children. The Verbal and Performance IQ scores of earlier versions of the scale have been replaced by four Composite/Index scores on the WISC-IV: Verbal Comprehension, Perceptual Reasoning, Working Memory and Processing Speed. The weight of *processing skills* in the Full Scale IQ calculation has doubled, with a consequent reduction in the weight assigned to *reasoning* tasks (verbal, visual-spatial and mathematical). Testers of the gifted know that abstract reasoning tasks best identify cognitive giftedness, while processing skills measures do not. Gifted children with or without disabilities may be painstaking, reflective and perfectionistic on paper-and-pencil tasks, lowering their Processing Speed Index scores; to a lesser degree, they may struggle when asked to recall non-meaningful material (Digit Span, Letter-Number Sequencing), lowering their Working Memory Index, even though they excel on meaningful auditory memory tasks that pique their interest.

As a result, a majority of gifted children show considerable variability in their Composite/Index scores on the WISC-IV, a problem less often encountered in average children. When this occurs, WISC-IV Full Scale IQ scores for the gifted may be difficult to interpret and, in some cases, may be lowered sufficiently by processing skills to prevent gifted children from qualifying for needed programs.

It is recommended practice to derive the General Ability Index (GAI) when there are large disparities among the Composite/Index scores (Flanagan & Kaufman, 2004; Weiss, Saklofske, Prifitera & Holdnack, 2006). Flanagan and Kaufman (2004), in *Essentials of WISC-IV Assessment*, deem the FSIQ "not interpretable" if Composite scores vary by 23 points (1.5 standard deviations) or more. The GAI utilizes only scores from the Verbal Comprehension and Perceptual Reasoning Composites, not Working Memory and Processing Speed. If the Verbal Comprehension and Perceptual Reasoning Composite scores vary by less than 23 points, "the GAI may be calculated and interpreted as a reliable and valid estimate of a child's global intellectual ability" (p. 128). Use of the GAI takes on special significance with the gifted. Verbal Comprehension and Perceptual Reasoning tasks are heavily loaded on abstract reasoning ability and are better indicators of giftedness than Working Memory (auditory memory that is manipulated) and Processing Speed (speed on paper-and-pencil tasks). Harcourt Assessments, publishers of the

WISC-IV, provides GAI tables on its website in support of similar use of the GAI when the variance between Composite scores is both *significant* and *unusual* (see *Technical Report #4*).

In light of these circumstances, where comprehensive testing is available, NAGC recommends that WISC-IV Full Scale IQ scores **not** be required for admission to gifted programs. Instead, the following guidelines are suggested:

> When the WISC-IV is used for the identification of gifted students, either the General Ability Index (GAI), which emphasizes reasoning ability, or the Full Scale IQ Score (FSIQ), should be acceptable for selection to gifted programs. The GAI should be derived using the table provided in the Harcourt Assessments website (*Technical Report 4*) [http://harcourtassessments.com/hai/Images/pdf/wisciv/WISCIVTechReport4.pdf]
>
> The Verbal Comprehension Index (VCI) and the Perceptual Reasoning Index (PRI) are also independently appropriate for selection to programs for the gifted, especially for culturally diverse, bilingual, twice exceptional students or visual-spatial learners. It is important that a good match be made between the strengths of the child and the attributes of the program. Students who have special learning needs should be admitted to gifted programs, provided that there are other indications of giftedness and instructional modifications are made to fit the needs of the students.

Testers should consider whether flexibility in subtest choice is needed. Up to two substitutions of supplementary subtests for core subtests can be made on the WISC-IV (in different Composite areas), decided a priori. For example, the use of Arithmetic, instead of Digit Span or Letter-Number Sequencing, may improve assessment of Working Memory for gifted children who are not math phobic. Arithmetic substitutes a meaningful memory task for one of the non-meaningful subtests, is heavily weighted for abstract reasoning ability, and can reveal mathematical talent. Substitutions may also be considered for disabilities, such as using Picture Completion instead of Block Design when testing a child with fine motor difficulties.

If these guidelines are followed, the WISC-IV offers an excellent reasoning test with a good balance between verbally administered abstract reasoning and language items and tasks that assess visual-spatial and nonverbal reasoning with visual prompts (minimal verbal explanation). Visual items on the WISC-IV offer reduced timing emphasis over those on the WISC-III, an advantage for reflective gifted children. The entire WISC-IV is a wise choice for the comprehensive assessment of gifted children, when Working Memory and Processing Speed subtests are used diagnostically. Administering just the Verbal Comprehension and Perceptual Reasoning sections (a total of six subtests), and calculating a GAI, is also a justifiable, shorter, and cost-effective alternative for selecting gifted students.

Selected References

Flanagan, D. P., & Kaufman, A. S. (2004). *Essentials of WISC-IV assessment.* Hoboken, NJ: John Wiley & Sons.

Gilman, B. J., & Falk, R. F. (2005, August). *Research-based guidelines for use of the WISC-IV in gifted assessment.* Paper presented at the 16th Biennial Conference of the World Council for Gifted and Talented Children, New Orleans, LA.

Lubin, B., Wallis, R. R., & Paine, C. (1971). Patterns of psychological test usage in the United States: 1935 – 1969. *Professional Psychology, 2,* 70-74.

Rimm, S., Gilman, B. J., & Silverman, L. K. (2008). Non-traditional applications of traditional testing. In J. VanTassel-Baska (Ed.), *Critical issues in equity and excellence in gifted education series, Volume 2: Alternative assessment of gifted learners* (pp. 175-202). Waco, TX: Prufrock Press.

Silverman, L. K. (in press). The measurement of giftedness. In L. Shavinina, Ed. *The handbook on giftedness.* New York: Springer Science.

Wechsler, D. (2003). *The WISC-IV technical and interpretive manual.* San Antonio, TX: Psychological Corporation.

Weiss, L. G., Saklofske, D. H., Prifitera, A., & Holdnack, J. A. (2006). *WISC-IV Advanced clinical interpretation.* Burlington, MA: Academic Press.

Approved 1/08

Norm- and Criterion-Referenced Testing

ERIC Digest
ERIC Identifier: ED410316
Publication Date: 1996-12-00
Author: Linda A. Bond

Tests can be categorized into two major groups: norm-referenced tests and criterion-referenced tests. These two tests differ in their intended purposes, the way in which content is selected, and the scoring process which defines how the test results must be interpreted. This brief paper will describe the differences between these two types of assessments and explain the most appropriate uses of each.

INTENDED PURPOSES

The major reason for using a norm-referenced tests (NRT) is to classify students. NRTs are designed to highlight achievement differences between and among students to produce a dependable rank order of students across a continuum of achievement from high achievers to low achievers (Stiggins, 1994). School systems might want to classify students in this way so that they can be properly placed in remedial or gifted programs. These types of tests are also used to help teachers select students for different ability level reading or mathematics instructional groups.

With norm-referenced tests, a representative group of students is given the test prior to its availability to the public. The scores of the students who take the test after publication are then compared to those of the norm group. Tests such as the California Achievement Test (CTB/McGraw-Hill), the Iowa Test of Basic Skills (Riverside), and the Metropolitan Achievement Test (Psychological Corporation) are normed using a national sample of students. Because norming a test is such an elaborate and expensive process, the norms are typically used by test publishers for 7 years. All students who take the test during that seven year period have their scores compared to the original norm group. While norm-referenced tests ascertains the rank of students, criterion-referenced tests (CRTs) determine "...what test takers can do and what they know, not how they compare to others (Anastasi, 1988, p. 102). CRTs report how well students are doing relative to a pre-determined performance level on a specified set of educational goals or outcomes included in the school, district, or state curriculum.

Educators or policy makers may choose to use a CRT when they wish to see how well students have learned the knowledge and skills which they are expected to have mastered. This information may be used as one piece of information to determine how well the student is learning the desired curriculum and how well the school is teaching that curriculum.

Both NRTs and CRTs can be standardized. The U.S. Congress, Office of Technology Assessment (1992) defines a standardized test as one that uses uniform procedures for administration and scoring in order to assure that the results from different people are comparable. Any kind of test--from multiple choice to essays to oral examinations--can be standardized if uniform scoring and administration are used (p. 165). This means that the comparison of student scores is possible. Thus, it can be assumed that two students who receive the identical scores on the same standardized test demonstrate corresponding levels of performance. Most national, state and district tests are standardized so that every score can be interpreted in a uniform manner for all students and schools.

SELECTION OF TEST CONTENT

Test content is an important factor choosing between an NRT test and a CRT test. The content of an NRT test is

selected according to how well it ranks students from high achievers to low. The content of a CRT test is determined by how well it matches the learning outcomes deemed most important. Although no test can measure everything of importance, the content selected for the CRT is selected on the basis of its significance in the curriculum while that of the NRT is chosen by how well it discriminates among students.

Any national, state or district test communicates to the public the skills that students should have acquired as well as the levels of student performance that are considered satisfactory. Therefore, education officials at any level should carefully consider content of the test which is selected or developed. Because of the importance placed upon high scores, the content of a standardized test can be very influential in the development of a school's curriculum and standards of excellence.

NRTs have come under attack recently because they traditionally have purportedly focused on low level, basic skills. This emphasis is in direct contrast to the recommendations made by the latest research on teaching and learning which calls for educators to stress the acquisition of conceptual understanding as well as the application of skills. The National Council of Teachers of Mathematics (NCTM) has been particularly vocal about this concern. In an NCTM publication (1991), Romberg (1989) cited that "a recent study of the six most commonly used commercial achievement tests found that at grade 8, on average, only 1 percent of the items were problem solving while 77 percent were computation or estimation" (p. 8).

In order to best prepare their students for the standardized achievement tests, teachers usually devote much time to teaching the information which is found on the standardized tests. This is particularly true if the standardized tests are also used to measure an educator's teaching ability. The result of this pressure placed upon teachers for their students to perform well on these tests has resulted in an emphasis on low level skills in the classroom (Corbett & Wilson, 1991). With curriculum specialists and educational policy makers alike calling for more attention to higher level skills, these tests may be driving classroom practice in the opposite direction of educational reform.

TEST INTERPRETATION

As mentioned earlier, a student's performance on an NRT is interpreted in relation to the performance of a large group of similar students who took the test when it was first normed. For example, if a student receives a percentile rank score on the total test of 34, this means that he or she performed as well or better than 34% of the students in the norm group. This type of information can useful for deciding whether or not students need remedial assistance or is a candidate for a gifted program. However, the score gives little information about what the student actually knows or can do. The validity of the score in these decision processes depends on whether or not the content of the NRT matches the knowledge and skills expected of the students in that particular school system.

It is easier to ensure the match to expected skills with a CRT. CRTs give detailed information about how well a student has performed on each of the educational goals or outcomes included on that test. For instance, "... a CRT score might describe which arithmetic operations a student can perform or the level of reading difficulty he or she can comprehend" (U.S. Congress, OTA, 1992, p. 170). As long as the content of the test matches the content that is considered important to learn, the CRT gives the student, the teacher, and the parent more information about how much of the valued content has been learned than an NRT.

SUMMARY

Public demands for accountability, and consequently for high standardized tests scores, are not going to disappear. In 1994, thirty-one states administered NRTs, while thirty-three states administered CRTs. Among these states, twenty-two administered both. Only two states rely on NRTs exclusively, while one state relies exclusively on a CRT. Acknowledging the recommendations for educational reform and the popularity of standardized tests, some

states are designing tests that "reflect, insofar as possible, what we believe to be appropriate educational practice" (NCTM, 1991, p.9). In addition to this, most states also administer other forms of assessment such as a writing sample, some form of open-ended performance assessment or a portfolio (CCSSO/NCREL, 1994).

Before a state can choose what type of standardized test to use, the state education officials will have to consider if that test meets three standards. These criteria are whether the assessment strategy(ies) of a particular test matches the state's educational goals, addresses the content the state wishes to assess, and allows the kinds of interpretations state education officials wish to make about student performance. Once they have determined these three things, the task of choosing between the NRT and CRT will becomes easier.

REFERENCES

Anastasi, A. (1988). Psychological Testing. New York, New York: MacMillan Publishing Company.

Romberg, T.A., Wilson, L. & Mamphono Khaketla (1991). "The Alignment of Six Standardized Tests with NCTM Standards", an unpublished paper, University of Wisconsin-Madison. In Jean Kerr Stenmark (ed; 1991). Mathematics Assessment: Myths, Models, Good Questions, and Practical Suggestions. The National Council of Teachers of Mathematics (NCTM)

Stenmark, J.K (ed; 1991). Mathematics Assessment: Myths, Models, Good Questions, and Practical Suggestions. Edited by. Reston, Virginia: The National Council of Teachers of Mathematics (NCTM)

Stiggins, R.J. (1994). Student-Centered Classroom Assessment. New York: Merrill

U.S. Congress, Office of Technology Assessment (1992). Testing in America's Schools: Asking the Right Questions. OTA-SET-519 (Washington, D.C.: U.S. Government Printing Office)

Connecting Performance Assessment to Instruction: A Comparison of Behavioral Assessment, Mastery Learning, Curriculum-Based Measurement, and Performance Assessment

ERIC Digest #E530
ERIC Identifier: ED381984
Publication Date: 1995-06-00
Author: Lynn S. Fuchs

A major impetus for the performance assessment movement has been the need to reconnect large-scale and classroom assessment to learning so that assessment affects learning positively, enhancing instruction.

IN WHAT WAYS CAN ASSESSMENT ENHANCE INSTRUCTION?

When teachers are better informed of the learning progress and difficulties of their students, they can make better decisions about what a student needs to learn next and how to teach that material in a manner that will maximize the student's learning. Teachers make three types of decisions using assessment results:

1. Instructional placement decisions--what the student knows and where he or she should be in the instructional sequence--i.e., what to teach next.
2. Formative evaluation decisions--information to monitor a student's learning while an instructional program is underway--how quickly progress is being made, whether the instructional program is effective, and whether a change in instructional program is needed to promote the student's learning.
3. Diagnostic decisions--which specific difficulties account for the student's inadequate progress so the teacher can remediate learning progress and design more effective instructional plans.

WHAT CRITERIA SHOULD ASSESSMENTS MEET IF THEY ARE TO INFORM INSTRUCTIONAL DECISIONS?

These assessments should meet seven criteria:

1. Measure important learning outcomes.
2. Address all three purposes of assessment.
3. Provide clear descriptions of student performance that can be linked to instructional actions.
4. Be compatible with a variety of instructional models.
5. Be easily administered, scored, and interpreted by teachers.
6. Communicate the goals of learning to teachers and students.
7. Generate accurate, meaningful information (i.e., be reliable and valid).

HOW DOES PERFORMANCE ASSESSMENT COMPARE TO OTHER METHODS OF LINKING ASSESSMENT TO INSTRUCTION?

Other methods of linking assessment to instruction include behavioral assessment, mastery learning, and curriculum-based measurement.

***Behavioral assessment**. Behavioral assessment relies on direct observation and recording of target behaviors, using repeated observations in the setting where the behavior occurs. Environmental factors (i.e., the situations in which the behaviors occur) and their effect on the behaviors are examined.

For example, if a teacher wanted to instruct a student in grocery shopping, she would first analyze the tasks associated with grocery shopping, put them in order, and design behavioral objectives that measure each task. Tasks

might include creating a shopping list, finding the items in the store, and finding the price of each item. The teacher would then collect data on each task to identify those in which the student needed instruction. The teacher would begin instruction at the point in the task sequence where the student was unable to correctly complete the task. Once the student could correctly complete a task, the teacher would move on to the next step, moving through the sequence until all of the tasks were mastered.

Behavioral assessment meets some but not all of the criteria for assessments listed above. It can inform the teacher about the student's placement in the instructional sequence and can help the teacher reach formative evaluation and diagnostic decisions. It communicates clearly what the essential learning content is, and it is feasible to administer, score, and interpret. In addition, its repeated measurements support the reliability of assessments. However, behavioral assessment tends to focus on discrete tasks that do not necessarily add up to important outcomes. It is limited to observable behaviors, and its small units of instruction can be difficult for students to piece together and apply to real-world outcomes. Additionally, the assessment system dictates a behavioral approach to instruction, which can limit the teacher's instructional options.

*Mastery learning.
In mastery learning, a curriculum is broken down into a set of subskills, which are then ordered in a hierarchy of instructional objectives. For each step in the instructional hierarchy, a criterion-referenced test is designed, and a performance criterion indicating mastery of the subskill is specified. The teacher starts at the lowest step in the hierarchy, pretests, teaches the objective, and posttests on the material. If the student does not demonstrate mastery, the teacher uses corrective strategies until mastery is achieved. The teacher then advances the student to the next, more difficult step in the hierarchy.

Like behavioral assessment, mastery learning provides information for instructional placement, formative evaluation, and diagnostic decisions. It communicates clearly to teachers and students about what is important to teach and learn. However, mastery learning suffers from the same limitation as behavioral assessment: it focuses on discrete behaviors in both assessment and instruction. Because little emphasis has been placed on its reliability or validity, users do not know what exactly is being assessed, how to interpret the resulting information, and how to use the measures effectively. Moreover, the measurement system dictates a specific approach to instruction, leaving the teacher few instructional choices. The focus of measurement changes each time a student achieves mastery of a step in the curriculum, and the steps may be of unequal difficulty, so progress cannot be judged over time. Finally, because different students need to be measured simultaneously on different steps of the curriculum, mastery learning systems can become unmanageable for teachers.

***Curriculum-based measurement (CBM).**
The focus of CBM is long-term. The teacher establishes a broad outcome for the student such as competently performing mathematics at the third-grade level at the end of the school year. Then the teacher uses CBM methods to measure student proficiency: he or she creates a pool of equivalent assessments, each of which samples the key problem types from the third grade curriculum. Each week, the student completes one or two assessments. Because each assessment is of equal difficulty and incorporates all of the important problem types to be learned over the year, the CBM data base produces a total score graphed over time to show progress over the year. Analysis of the student's performance on separate skills embedded in the assessment can also be conducted for diagnostic problem-solving to improve the instructional program.

CBM satisfies six of the criteria for assessments. It addresses the three purposes for assessment, and it incorporates standardized measurement techniques, providing reliability and validity. It offers detailed information on a student's performance on specific skills and can be used to determine how to improve an instructional program. Its

measurement framework is not tied to any particular model of instruction, so a broad range of instructional options can be used. A teacher can use widely varying methods with the same child to see which method is most beneficial. Students know how they are evaluated and can set personal learning goals. In addition, the assessment demands are manageable in classroom settings, and to make them even more easily manageable, computer programs have been developed to administer assessments and manage the data.

However, CBM has two drawbacks with respect to the criteria for assessments. The system requires longer time periods to reveal growth, and the connection between assessment results and instructional decisions is not as clear as with behavioral assessment or mastery learning. Controversy also exists about the importance of the learning outcomes associated with CBM. That is, it relies on pencil and paper tasks in math and spelling and one-dimensional assessments in reading, while current discussions about outcomes stress the utility of multidimensional measures that can cut across curriculum areas.

***Performance assessment.**

Three key features of performance assessment are: (1) students construct, rather than select, responses; (2)assessment formats allow teachers to observe student behavior on tasks reflecting real-world requirements; and (3) scoring reveals patterns in students' learning and thinking.

An example of a performance assessment task is provided below:

> A group of five families on your block is going to have a garage sale in which clothes, toys, and books will be sold. Your family has 12 items to sell and will need 18 square feet to display these items; the Hamletts have 13 items and need 20 square feet; the Phillips, 7 items and 10 square feet; the Garcias, 15 items and 15 square feet; the Nguyens, 10 items and 30 square feet. Rental tables measure 6 feet by 2.5 feet and cost $6.00 a day. The garage where the sale will be held is 20 feet by 30 feet. Newspaper advertising costs $11 for the first 10 words and $1.50 for each additional word.
>
> 1. How many tables will you need? Explain how you got this number.
> 2. Draw a diagram showing how the tables can be arranged in the garage to allow the customers to move about with at least 4 feet between tables.
> 3. Write an ad for your sale that includes enough information.
> 4. How much money do you have to earn from your sale for the families to break even?

The students are aware of the scoring system and the criteria used to determine the scores. Their responses will be classified as exemplary, competent, minimal, inadequate, or no attempt based on a rubric that specifies the characteristics of responses in each of these categories. This problem offers one version of what a teacher's use of performance assessment might look like. In practice, many varieties of performance assessment are used. This problem measures massed mathematical concepts that include addition, multiplication, decimals, data analysis, perimeters, area, spatial sense, graphic representation, money, and communication about mathematics. Students take about 50 minutes for the assessment, and it can be completed individually or in small groups. The problem is anchored in a real-life, age-appropriate situation and represents real applications of mathematics.

HOW WELL DOES PERFORMANCE ASSESSMENT SATISFY THE SEVEN CRITERIA FOR ASSESSMENT?

Today, performance assessment is relatively new, undeveloped, and yet to be studied systematically. Many practitioners are experimenting with its use and contributing to its development and refinement. Yet they are often in the undesirable position of interpreting vague design features and operationalizing those features into specific assessments on their own. These assessments take a variety of forms, some of which are closer than others in approximating the conceptual and theoretical underpinnings of performance assessment.

1. Measure important learning outcomes. The extent to which performance assessment measures important student outcomes depends on the specific assessment problem or task. Performance assessment tasks should reflect important, real-world performances that are tied to desired student outcomes that are relevant to the workplace and everyday life. They should connect meaningfully with specific instructional methods that can be realistically managed in school settings.

2. Address all three purposes of assessment. It is unclear how performance assessment can be used to formulate instructional placement or formative evaluation decisions. Ideally, alternate forms of the problem could include the same concepts administered over time in order to yield information about individual students' progress. Although performance assessment offers the promise of addressing all three assessment purposes, specific methods for doing so have yet to be developed.

3. Provide clear descriptions of student performance that can be linked to instructional actions. When performance assessment tasks address a variety of concepts in age-appropriate, real-world situations, teachers can form a picture of student performance across skills and identify the student's problem-solving strategies. However, this depends on the teacher's skill in identifying student competencies, gleaning information about students' strategic behavior, and relating these observations to specific instructional techniques. Consultation methods or computerized strategies for generating profiles of student competence are needed.

4. Be compatible with a variety of instructional models. Theoretically, performance assessment could be used with a variety of instructional approaches. Teachers should experiment with a variety of instructional methods as they implement performance assessment, especially with students who have serious learning problems.

5. Be easily administered, scored, and interpreted by teachers. Performance assessment can require large amounts of teacher time to design and administer assessments and to scrutinize student performances. It is easy to see how this type of assessment could generate so many different plans for intervention strategies for different students that teachers in a classroom situation with 20 or 30 students would be unable to manage. Performance assessment developers need to solve the problem of how to implement plans based on performance assessments within the constraints of classroom life.

6. Communicate the goals of learning to teachers and students. When it is clearly apparent that an assessment is aligned with instructional goals, teachers should be able to use that assessment to direct their instruction, and students should be able to use it to establish personal learning goals. This depends, however, on the extent to which the scoring rubric used is clear, concrete, and visible.

7. Generate accurate, meaningful information (i.e., be reliable and valid). Performance assessment represents a vision that can shape the future direction of classroom-based assessment, but it requires much additional scrutiny and development before it can fulfill its promise.

REFERENCES

Derived from Fuchs, L. (1994). Connecting Performance Assessment to Instruction. Reston, VA: The Council for Exceptional Children. (Product #P5058).

Creating Meaningful Performance Assessments

ERIC Digest #E531
ERIC Identifier: ED381985
Publication Date: 1995-06-00
Author: Stephen N. Elliott

Performance assessment is a viable alternative to norm-referenced tests. Teachers can use performance assessment to obtain a much richer and more complete picture of what students know and are able to do.

DEFINING PERFORMANCE ASSESSMENT

Defined by the U.S. Congress, Office of Technology Assessment (OTA) (1992), as "testing methods that require students to create an answer or product that demonstrates their knowledge and skills," performance assessment can take many forms including:

*Conducting experiments.
*Writing extended essays.
*Doing mathematical computations.

Performance assessment is best understood as a continuum of assessment formats ranging from the simplest student-constructed responses to comprehensive demonstrations or collections of work over time. Whatever format, common features of performance assessment involve:

1. Students' construction rather than selection of a response.
2. Direct observation of student behavior on tasks resembling those commonly required for functioning in the world outside school.
3. Illumination of students' learning and thinking processes along with their answers (OTA, 1992).

Performance assessments measure what is taught in the curriculum. There are two terms that are core to depicting performance assessment:

1. Performance: A student's active generation of a response that is observable either directly or indirectly via a permanent product.
2. Authentic: The nature of the task and context in which the assessment occurs is relevant and represents "real world" problems or issues.

HOW DO YOU ADDRESS VALIDITY IN PERFORMANCE ASSESSMENTS?

The validity of an assessment depends on the degree to which the interpretations and uses of assessment results are supported by empirical evidence and logical analysis. According to Baker and her associates (1993), there are five internal characteristics that valid performance assessments should exhibit:

1. Have meaning for students and teachers and motivate high performance.
2. Require the demonstration of complex cognition, applicable to important problem areas.
3. Exemplify current standards of content or subject matter quality.
4. Minimize the effects of ancillary skills that are irrelevant to the focus of assessment.
5. Possess explicit standards for rating or judgment.

When considering the validity of a performance test, it is important to first consider how the test or instrument "behaves" given the content covered. Questions should be asked such as:

*How does this test relate to other measures of a similar construct?

*Can the measure predict future performances?
*Does the assessment adequately cover the content domain?

It is also important to review the intended effects of using the assessment instrument. Questions about the use of a test typically focus on the test's ability to reliably differentiate individuals into groups and guide the methods teachers use to teach the subject matter covered by the test.

A word of caution: Unintended uses of assessments can have precarious effects. To prevent the misuse of assessments, the following questions should be considered:
*Does use of the instrument result in discriminatory practices against various groups of individuals?
*Is it used to evaluate others (e.g., parents or teachers) who are not directly assessed by the test?

PROVIDING EVIDENCE FOR THE RELIABILITY AND VALIDITY OF PERFORMANCE ASSESSMENT

The technical qualities and scoring procedures of performance assessments must meet high standards for reliability and validity. To ensure that sufficient evidence exists for a measure, the following four issues should be addressed:

1. Assessment as a Curriculum Event. Externally mandated assessments that bear little, if any, resemblance to subject area domain and pedagogy cannot provide a valid or reliable indication of what a student knows and is able to do. The assessment should reflect what is taught and how it is taught. Making an assessment a curriculum event means reconceptualizing it as a series of theoretically and practically coherent learning activities that are structured in such a way that they lead to a single predetermined end. When planning for assessment as a curriculum event, the following factors should be considered:
*The content of the instrument.
*The length of activities required to complete the assessment.
*The type of activities required to complete the assessment.
*The number of items in the assessment instrument.
*The scoring rubric.

2. Task Content Alignment with Curriculum. Content alignment between what is tested and what is taught is essential. What is taught should be linked to valued outcomes for students in the district.

3. Scoring and Subsequent Communications with Consumers. In large scale assessment systems, the scoring and interpretation of performance assessment instruments is akin to a criterion-referenced approach to testing. A student's performance is evaluated by a trained rater who compares the student's responses to multitrait descriptions of performances and then gives the student a single number corresponding to the description that best characterizes the performance. Students are compared directly to scoring criteria and only indirectly to each other.

In the classroom, every student needs feedback when the purpose of performance assessment is diagnosis and monitoring of student progress. Students can be shown how to assess their own performances when:
*The scoring criteria are well articulated.
*Teachers are comfortable with having students share in their own evaluation process.

4. Linking and Comparing Results Over Time. Linking is a generic term that includes a variety of approaches to making results of one assessment comparable to those of another. Two appropriate and manageable approaches to linking in performance assessment include:
*Statistical Moderation. This approach is used to compare performances across content areas for groups of

students who have taken a test at the same point in time.

*Social Moderation. This is a judgmental approach that is built on consensus of raters. The comparability of scores assigned depends substantially on the development of consensus among professionals.

HOW CAN TEACHERS INFLUENCE STUDENTS' PERFORMANCES?

Performance assessment is a promising method that is achievable in the classroom. In classrooms, teachers can use data gathered from performance assessment to guide instruction. Performance assessment should interact with instruction that precedes and follows an assessment task.

When using performance assessments, students' performances can be positively influenced by:

1. Selecting assessment tasks that are clearly aligned or connected to what has been taught.
2. Sharing the scoring criteria for the assessment task with students prior to working on the task.
3. Providing students with clear statements of standards and/or several models of acceptable performances before they attempt a task.
4. Encouraging students to complete self-assessments of their performances.
5. Interpreting students' performances by comparing them to standards that are developmentally appropriate, as well as to other students' performances.

REFERENCES

Baker, E. L., O'Neill, H. F., Jr., & Linn, R. L. (1993). Policy and validity prospects for performance-based assessments. American Psychologist, 48, 1210-1218.

U.S. Congress, Office of Technology Assessment. (1992, February). Testing in American schools: Asking the right questions. (OTA-SET-519). Washington, DC: U.S. Government Printing Office.

Derived from: Elliot, S. N. (1994). Creating Meaningful Performance Assessments: Fundamental Concepts. Reston, VA: The Council for Exceptional Children. Product #P5059.

Commonly Used Testing Instruments in Gifted Education, Updated 4/02/

Leah Ellis,
Formerly the Gifted and Talented Consultant
Division of Curriculum Development
Office of Teaching and Learning
Kentucky Department of Education

Following is a listing of some commonly used testing instruments to assess ability or screen for potential giftedness.

The tests and publishers included in this list do not necessarily reflect the policy or viewpoint of the Kentucky Department of Education, nor does the mention of a particular organization or product imply endorsement or compliance with educational regulations for gifted and/or assessment.

California Achievement Tests, Fifth Edition (CAT-5)
A traditional assessment series that provides comprehensive evaluation of student achievement in reading, language, spelling, math, study skills, science and social studies.
Grades: K-12
CTB McGraw-Hill
www.ctb.com

Creativity Assessment Packet (CAP)
This Frank Williams Creativity Test measures the cognitive thought factors of fluency, flexibility, elaboration, originality, vocabulary, and comprehension.
Ages 6-18
PRO-ED, Inc.
www.proedinc.com

Cognitive Abilities Test® (CogAT)
Group administered test battery to assess ability in reasoning and problem solving
Grades K-12
Riverside Publishing
www.riverpub.com

Comprehensive Test of Nonverbal Intelligence (CTONI)
Measures nonverbal reasoning abilities
Ages 6-90
PRO-ED, Inc.
www.proedinc.com

Das-Naglieri Cognitive Assessment System (CAS)
Evaluates cognitive processing
Ages 5 to 17
Riverside Publishing
www.riverpub.com

Detroit Tests of Learning Aptitude (DTLA-4)
Measures both general intelligence and discrete ability areas
Ages 6-17
PRO-ED, Inc.
www.proedinc.com

Differential Ability Scales (DAS)
Measures cognitive abilities and achievement.
Ages 2.6 to 17.11
Harcourt Assessment, Inc.
www.harcourtassessment.com

Draw-A-Person Intellectual Ability Test for Children, Adolescents, and Adults (DAP:IQ)
Estimates intellectual ability from human figure drawings
Ages 4 to 89
Western Psychological Services
www.wpspublish.com

Gifted and Talented Evaluation Scales (GATES)
A norm-referenced instrument that assesses the characteristics, skills, and talents of gifted students
Ages 5 to 18
PRO-ED, Inc.
www.proedinc.com

Gifted Evaluation Scale, Second Edition (GES-2)
Designed to help identify gifted students
Grades K to 12
Hawthorne Educational Services, Inc.
www.hes-inc.com/hes.cgi

Gifted Rating Scales
Norm-referenced rating scales designed to assess observable student behaviors indicating giftedness.
Grades: K-8
Harcourt Assessment Inc.
www.harcourtassessment.com

Gray Oral Reading Tests (GORT-4)
The *Gray Oral Reading Test* has been revised and all new normative data provided. All scores are reported in terms of standard scores, percentile ranks, grade equivalents, and age equivalents. The Fluency and Oral Reading Comprehension Score are combined to obtain an Oral Reading Quotient.
Age Range: Ages 6-18
PRO-ED, Inc.
www.proedinc.com

Group Reading Assessment and Diagnostic Evaluation (GRADE)
Reading assessment including: Percentile ranks, standard scores, grade equivalents, stanines
Age Range: Pre-kindergarten through adult
Pearson Learning
http://www.pearsonlearning.com/grade/index.cfm

Group Mathematics Assessment and Diagnostic Evaluation (G•MADE)
Math assessment including: Stanines, percentiles, grade and age equivalents, standard scores
Age Range: Kindergarten through adult
Pearson Learning
http://www.pearsonlearning.com/gmade/index.cfm

In View
An assessment of cognitive abilities that includes verbal reasoning, sequences, analogies, and quantitative reasoning. Provides academic ability scores and test scores for placement decisions in Gifted and Talented and other programs.
Grades 2-12
CTB McGraw-Hill
www.ctb.com

Iowa Acceleration Scale (2nd Edition, Complete Kit)
A Guide for Whole-Grade Acceleration
Grades K-8
Great Potential Press
http://www.giftedbooks.com/productdetails.asp?id=92

Iowa Algebra Aptitude Test, Fifth Edition (IAAT)
A group administered algebra placement test, assessing placement in Algebra I
Grades: 7 and 8 but suitable in lower grades for accelerated students, as well as high school.
Riverside Publishing
www.riverpub.com

Iowa Tests of Basic Skills® (ITBS®) Forms A and B
Group administered achievement test battery. Co-normed with the *Iowa Tests of Educational Development* ® and the *Cognitive Abilities Test* ™ and web-based reporting.
Grades: K-8 (Levels 5-14)
Riverside Publishing
www.riverpub.com/products/itbs/index.html

Kaufman Assessment Battery for Children, Second Edition (KABC-II)
A culturally fair, individually administered measure of processing and cognitive ability.
Ages 3 to 18
Pearson Assessments
www.pearsonassessments.com/

Kaufman Brief Intelligence Test, Second Edition (KBIT-2)
A recently revised, individually administered, quick measure of verbal and nonverbal cognitive ability
Ages 4 to 90
Pearson Assessments
www.pearsonassessments.com/

Kaufman Test of Educational Achievement, Second Edition (KTEA-II)
An individually administered battery in reading, math, written language, and oral language
Ages 4 to 90+
Pearson Assessments
www.pearsonassessments.com/

KeyMath 3
Assess critical math skills with greater integration and conceptual rigor. Several types of norm-referenced scores are provided, including scale and standard scores, percentile ranks, age and grade equivalents, and growth scores.
Grades: Pre K - 12
Pearson Assessments
www.pearsonassessments.com

Khatena-Morse Multi-talent Perception Inventory (KMMPI)
Identifies giftedness in music, art, and leadership
Ages: 6 – adult
Scholastic Testing Service, Inc.
www.ststesting.com

Khatena-Torrance Creative Perception Inventory (KTCPI)
Identifies candidates for creativity programs
Ages: 12- adult
Scholastic Testing Service, Inc.
www.ststesting.com

Kuhlmann-Anderson Tests, Eighth Edition (KA)
Measure academic potential through assessing cognitive skills related to the learning process
Grades: K-12
Scholastic Testing Service, Inc.
www.ststesting.com

Leadership Skills Inventory (LSI)
Assess strengths and weaknesses in the area of leadership
PRO-ED, Inc.
www.proedinc.com or www.crgleader.com

Leadership Strengths Indicator (LSI)
The LSI is a self-report leadership analysis instrument for adolescents with an extensive manual for initiating discussion about leadership style, component behaviors and attitudes of leadership and the ways in which specific leadership characteristics can be strengthened.
Grades 6-12
Royal Fireworks Press
http://www.rfwp.com/191X.htm

Leiter International Performance Scale-Revised (LEITER-R)
A nonverbal intelligence and cognitive assessment
Ages: 2-21
Stoelting Co.
www.stoeltingco.com

Matrix Analogies Test – Expanded Form (MAT)
Assesses nonverbal reasoning abilities
Ages 5 to 17
Harcourt Assessment, Inc.
www.harcourtassessment.com

Measures of Academic Progress (MAP)
State-aligned, norm-referenced, achievement tests that give a percentile score.
MAP tests are available in these areas: Mathematics, Reading, Language Usage, & Science.
http://www.nwea.org/assessments/map.asp

Metropolitan Achievement Tests®, Eighth Edition (MAT-8)
A complete battery that includes math, reading/language, science & social studies that measures foundational skills and critical thinking processes and strategies.
Grades K to 12
Harcourt Assessment, Inc
www.harcourtassessment.com

Naglieri Nonverbal Ability Test (NNAT)
Nonverbal group measure of reasoning and problem solving
Grades K to 12
Harcourt Assessment, Inc.
www.harcourtassessment.com

Otis-Lennon School Abilities Test, 8th edition (OLSAT)
Measures abstract thinking and reasoning ability
Grades K-12
Harcourt Assessment, Inc.
www.harcourtassessment.com

Peabody Individual Achievement Test (PIAT-R)
Measures academic achievement in reading math & spelling
Ages 5 to 22.11
Pearson Assessments
http://www.pearsonassessments.com/

Primary Measures of Music Audiation (K-3)
Intermediate Measures of Music Audiation (Grades 1-6)
Author: Edwin E. Gordon. Two aptitude tests designed to diagnose and measure music potential.
Grades K – 6
GIA Publications, Inc.
www.giamusic.com/scstore/P-musicaudiation.html

Primary Test of Cognitive Skills (PTCS)
A group-administered, intellectual functioning or ability test that assesses cognitive abilities including memory and verbal reasoning to identify giftedness, learning disabilities, or unique developmental delays. Can also assist with identifying anticipated achievement when used in conjunction with achievement tests.
Grades K-1
CTB McGraw-Hill
www.ctb.com

Process Skills Ratings Scales (PSRS)
Designed for use in obtaining a rating of the student's facility in using process skills that develop ability to think, reason, and search for knowledge independently and to communicate and interact effectively with all members of society.
PRO-ED, Inc.
www.proedinc.com

Raven's Progressive Matrices
Nonverbal assessment of perception and thinking Skills
Ages 5 to adult
Harcourt Assessment Inc.
www.harcourtassessment.com

Reynolds Intellectual Assessment Scales (RIAS)
Assesses verbal and nonverbal intelligence and memory
Ages 3 to 94
Psychological Assessment Resources, Inc
www3.parinc.com

Scales for Identifying Gifted Students (SIGS)
This standardized, norm-referenced instrument is completed by teachers or parents and provides a method for identifying gifted children.
Ages 5 to 18
PRO-ED, Inc.
www.proedinc.com

Scales for Rating the Behavioral Characteristics of Superior Students
Designed to aid teachers in the identification of gifted children
Grades K-12
Creative Learning Press, Inc.
www.creativelearningpress.com

Screening Assessment for Gifted Elementary & Middle School Students (SAGES-2)
Used to identify students who are gifted in academics and reasoning
Ages 5 to 14.11
PRO-ED, Inc.
www.proedinc.com

Slosson Full Range Intelligence Test (S-FRIT)
A brief individual screen of verbal/nonverbal intelligence, giving a balanced measure of cognitive ability, allowing an IQ range of 35-164
Ages: 5 – adult
Slosson Educational Publications
www.slosson.com

Slosson Intelligence Test – Primary (SIT-P)

Screening test of children's intelligence, giving a balanced measure of a child's cognitive ability, allowing an IQ range of 10-170+
Ages: 2 – 7.11
Slosson Educational Publications
www.slosson.com

Slosson Intelligence Test – Revised (SIT-R)
Estimates general verbal cognitive ability
Ages 4 to 65
Slosson Educational Publications
www.slosson.com

Stanford-Binet Intelligence Scales – 5th Edition (SB5)
A widely used individual intelligence test assessing cognitive abilities and development. The final composite score is viewed as a "global' measurement of cognitive ability.
Ages 2-85+
Western Psychological Services & Riverside Publishing
www.assess.nelson.com/test-ind/stan-b5.html & www.riverpub.com
http://www.amendpsych.com/stanfordbinet.html

Structure of Intellect Learning Abilities Test (SOI-LA)
Assesses a wide variety of cognitive abilities or factors of intelligence
Grades 2 to adult
Western Psychological Services
www.wpspublish.com

TerraNova CTBS
CTBS Multiple assessments assess Reading/Language Arts, Math, Science, Social Studies
Grades: 1-12
CTB McGraw-Hill
www.ctb.com

Test of Cognitive Skills, Second Edition (TCS/2)
An assessment of academic aptitude that includes verbal, nonverbal, and memory skills.
Grades 2-12
CTB McGraw-Hill
www.ctb.com

Test of Early Mathematical Ability, Third Edition (TEMA-3)
A norm-referenced measure that can identify gifted students in math.
Ages 3 through 8
PRO-ED, Inc.
www.proedinc.com

Tests of Mathematical Abilities (TOMA-2)
A norm-referenced test measuring math performance.
Grades 3 -12
PRO-ED, Inc.
www.proedinc.com

Tests of Mathematical Abilities for Gifted Students (TOMAGS)
Designed to identify students who have talent or giftedness in mathematics
Grades K to 6
PRO-ED, Inc.
www.proedinc.com

Tests of Nonverbal Intelligence (TONI-3)
Non-verbal measure of intelligence, aptitude, abstract reasoning, and problem solving.
Ages 6 to 89
PRO-ED, Inc.
www.proedinc.com

The ACT®
The ACT® test is a nationally administered, standardized test that helps colleges evaluate candidates. It assesses high school students' general educational development and their ability to complete college-level work. The multiple-choice tests cover four skill areas: English, mathematics, reading, and science.
Grades: 6-12
www.act.org

The SAT® Reasoning Test

The SAT® is a standardized measure of a student's college readiness, measuring critical reading, writing, and mathematical reasoning skills.
Grades: 6-12
www.collegeboard.com

Thinking Creatively in Action and Movement (TCAM)
Assesses the creativity of young children or others with limited verbal and drawing skills.
Ages: 3-6
Scholastic Testing Service, Inc.
www.ststesting.com

Thinking Creatively with Sounds and Words (TSCW)
Assesses creativity.
Grades: 3-12
Scholastic Testing Service, Inc.
www.ststesting.com

Torrance Tests of Creative Thinking- Figural (TTCT-F)
Identifies and evaluates creative potential and measure creative thinking
Ages: Kindergarten to adult
Scholastic Testing Service, Inc.
www.ststesting.com

Torrance Tests of Creative Thinking- Verbal
Identifies and evaluates creative potential and measure creative thinking
Ages: First Graders to adult
Scholastic Testing Service, Inc.
www.ststesting.com

Universal Nonverbal Intelligence Test (UNIT)
Measure of the general intelligence and cognitive abilities
Ages 5-17
Riverside Publishing
www.riverpub.com

Wechsler Individual Achievement Test - Second Edition (WIAT-II)
A comprehensive measurement tool useful for achievement skills assessment, learning disability diagnosis, special education placement, curriculum planning, and clinical appraisal for preschool children through adults.
Ages 4-85
Harcourt Assessment Inc.
www.harcourtassessment.com

Wechsler Intelligence Scale for Children (WISC-IV)
Assesses the cognitive ability of children
Ages 6 to 16.11
Harcourt Assessment Inc.
www.harcourtassessment.com

Wechsler Preschool and Primary Scale of Intelligence – Third Edition (WPPSI-III)
Assesses the intelligence for young children
Ages 2.6 to 7.3
Harcourt Assessment Inc.
www.harcourtassessment.com

Wide Range Achievement Test 3 (WRAT3)
Measures the development of reading, spelling, and arithmetic skills
Ages 4 to 85
Psychological Assessment Resources, Inc.
http://www3.parinc.com

Woodcock Johnson III
Contains two distinct, co-normed batteries that measure general intellectual ability, specific cognitive abilities, scholastic aptitude, oral language, and academic achievement
Ages 2 to 90+
Riverside Publishing
www.riverpub.com

Additional Testing Links
American Psychologocal Association (APA)
www.apa.org/science/testing.html

Buros Mental Measurements Yearbook
http://buros.unl.edu/buros/jsp/search.jsp

Educational Testing Service's database
www.ets.org/testcoll

Hoagies Gifted Testing & Assessment
http://www.hoagiesgifted.org/tests.htm#sb
http://www.hoagiesgifted.org/testing.htm

Additional Music testing from GIA Publications, Inc.
For measuring music potential and testing acquired knowledge.
http://www.giamusic.com/music_education/Music-Testing.html

How to Write a Test for both Creativity and Knowledge
http://www.goshen.edu/~marvinpb/arted/testing/drawtest.html

: