Macmillan Social Science Library

PSYCHOLOGY *of* CLASSROOM LEARNING

An Encyclopedia

ERIC M. ANDERMAN, *Editor in Chief*
LYNLEY H. ANDERMAN, *Co-editor*

Madaus, G. F., Lynch, C. A., & Lynch, P. S. (2001). A brief history of attempts to monitor testing. *National Board of Educational Testing and Public Policy Statements, 2*(2). Boston, MA: NBETPP.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). Washington, DC: American Council on Education-Macmillan.

National Association of Collegiate Admissions Counseling. (2001). *Statement of Principles of Good Practice.* Alexandria, VA: Author.

National Association of School Psychologists. (1997). *Procedural guidelines for the adjudications of ethical complaints.* Bethesda, MD: Author.

National Association of School Psychologists. (2000). *Professional Conduct Manual* (4th ed.). Bethesda, MD: Author.

National Council of Measurement in Education. (1995). *Code of professional responsibilities in education.* Washington, DC: Author.

Novick, M. R. (1981). Federal guidelines and professional standards. *American Psychologist, 36,* 1035–1046.

Phillips, S. E., & Camara, W. J. (2006). Legal and ethical issues. In Brennan, R. (Ed.), *Educational measurement* (ACE/Praeger Series on Higher Education) (4th ed., pp. 734–755). Westport, CT: Praeger.

Rhoades, K., & Madaus, G. (2003, May). *Errors in standardized testing: A systematic problem.* Retrieved November 1, 2007, from http://www.bc.edu/research/nbetpp/statements/M1N4.pdf.

Schmeiser, C. B. (1992). Ethical codes in the professions. *Educational Measurement: Issues and Practice, 11*(3), 5–11.

Wise, L. (2006). Encouraging and supporting compliance with standards for educational testing. *Educational Measurement: Issues and Practice, 25*(3), 51–56.

Wayne J. Camara

STANFORD-BINET INTELLIGENCE SCALES

The Stanford-Binet Intelligence Scales: Fifth Edition (SB-5; Roid, 2003a) is a test of intelligence/cognitive abilities for individuals 2 to over 85 years of age (child, adolescent, and adult). It is a major revision of the Stanford-Binet Intelligence Scales: Fourth Edition (SB-4; Thorndike, Hagen, & Sattler, 1986) and took seven years to complete. The complete SB-5 takes between 45 and 75 minutes to administer while the Abbreviated Battery takes between 15 and 20 minutes to complete. The Abbreviated Battery was included to allow for quick estimation of general intellectual abilities for screening purposes and includes the Nonverbal Fluid Reasoning and Verbal Knowledge subtests. These subtests are also used as routing subtests to provide estimates of intellectual functioning for placement into the appropriate level of test items better matching the individual's abilities. Use of the SB-5 may be for assessing mental retardation, learning disabilities, developmental disabilities, and intellectual giftedness, although diagnosis of mental retardation also requires assessment of and significant deficits in adaptive behaviors. The SB-5 examiner's manual articulates test user qualifications, including college and/or graduate training in statistics and measurement for understanding test scores; thorough understanding of standardized administration, scoring, and calculation of standardized scores; and supervised training in administration via training workshops and/or graduate school testing courses.

HISTORICAL BACKGROUND AND DEVELOPMENT

The SB-5 evolved out of the original pioneering work of Alfred Binet (1857–1911), Victor Henri, and Théodore Simon in France during the late 1800s and early 1900s. Binet and Henri (1895) defined intelligence in terms of complex mental abilities (i.e., memory, abstraction, judgment, and reasoning) whereas Sir Francis Galton (1822–1911) had previously relied primarily on measuring physical and sensory abilities in assessing individual differences. Binet and Henri also developed tasks to measure these complex mental abilities. In 1904 the Minister of Public Instruction in Paris established a committee charged with finding a means to differentiate mentally retarded from normal children, and Binet was appointed to this committee. Earlier work by Binet and Henri and collaboration with Simon evolved into the first practical test of intellectual abilities for the diagnosis of mental retardation: the Binet-Simon Scale of Intelligence (Binet & Simon, 1905). Revisions and improvements of the Binet-Simon Scale of Intelligence were published in 1908 and 1911. Due to the success of the Binet-Simon Scale in France; Henry H. Goddard, Frederick Kuhlmann, J. E. Wallace Wallin, and Robert M. Yerkes each created translations of the Binet-Simon Scale of Intelligence for their use in the United States (Kaufman, 1990), but these different translations were not comparable and proved problematic. It was Lewis M. Termin who provided the most comprehensive translation and adaptation of the Binet-Simon Scale of Intelligence and provided better standardization. Termin's measure has survived to the present day through numerous revisions.

STANFORD-BINET INTELLIGENCE SCALES-FIFTH EDITION (SB-5)

The Stanford-Binet Intelligence Scales: Fifth Edition, (SB-5; Roid, 2003a), is a major revision and restructuring based on the hierarchical model of intelligence measurement illustrated by John B. Carroll (1993) and previous work by Raymond B. Cattell (1943, 1963) and John L. Horn (Cattell & Horn, 1978). The Cattell-Horn-Carroll

Stanford-Binet Intelligence Scales: Fifth Edition subtests and activities in relation to verbal and nonverbal domains and CHC Stratum II Factors

	Domains	
Factors	**Nonverbal**	**Verbal**
Fluid reasoning	**Nonverbal fluid reasoning** *Object series/matrices**	**Verbal fluid reasoning** *Early reasoning* (Levels 2–3) *Verbal absurdities* (Level 4) *Verbal analogies* (Levels 5–6)
Knowledge	**Nonverbal knowledge** *Procedural knowledge* (Levels 2–3) *Picture absurdities* (Levels 4–6)	**Verbal knowledge** *Vocabulary**
Quantitative reasoning	**Nonverbal quantitative reasoning** *Quantitative reasoning* (Levels 2–6)	**Verbal quantitative reasoning** *Quantitative reasoning* (Levels 2–6)
Visual-spatial processing	**Nonverbal visual-spatial processing** *Form board* (Levels 1–2) *Form patterns* (Levels 3–6)	**Verbal visual-spatial processing** *Position and direction* (Levels 2–6)
Working memory	**Nonverbal working memory** *Delayed response* (Level 1) *Block span* (Levels 2–6)	**Verbal working memory** *Memory for sentences* (Levels 2–3) *Last word* (Levels 4–6)

Note: *Routing subtests. Subtests presented in bold. Table adapted from Roid (2003b) Figure 2.1.

Table 1 ILLUSTRATION BY GGS INFORMATION SERVICES. CENGAGE LEARNING, GALE.

(CHC) model of the structure of intellectual abilities is hierarchical with 50 to 60 narrow abilities at the bottom (Stratum I), 8 to 10 broad ability factors in the middle (Stratum II), and the general (*g*) ability factor at the top (Stratum III). The SB-5 activities and subtests measure a number of Stratum I dimensions, SB-5 Factors measure five Stratum II dimensions, and the Full Scale IQ measures Stratum III (*g* [general intelligence]). A number of subtests from the SB-4 were eliminated while new subtests were created and included in the SB-5.

The SB-5 (Roid, 2003a) includes ten subtests selected and designed to measure five CHC factors (fluid reasoning, knowledge, quantitative reasoning, visual-spatial processing, and working memory) within verbal and nonverbal domains. A global, Full Scale IQ score is provided in addition to Verbal IQ, Nonverbal IQ, and five composite factor scores. All scores are based on a mean of 100 and standard deviation of 15. Table 1 presents the SB-5 subtests (in bold) and the activities within subtests (in italics) used to measure the subtests specific to the different levels. Performance on the two routing subtests (nonverbal fluid reasoning and verbal knowledge) place individuals into the level appropriate for assessment. Recommendations for interpretation of SB-5 scores include the Full Scale IQ, comparisons of the Verbal and Nonverbal IQs, and the five factor scores (Roid, 2003b, 2003c).

Full Scale IQ scores range from 40 to 190, covering a wide range of intellectual abilities (±4 *SD*s). This allows for assessment to the lower levels of moderate mental retardation to the higher levels of intellectual giftedness. Verbal IQs range from 43 to 156 and Nonverbal IQs range from 42 to 158, providing a wide range. Factor scores also have wide ranges of possible scores (fluid reasoning: 47–153, knowledge: 49–151, quantitative reasoning: 50–149, visual–spatial processing: 48–152, working memory: 48–152).

The standardization sample of the SB-5 was stratified to closely match the 1998 United States Census data on key demographic variables of geographic region, race/ethnicity, age, and socioeconomic level for generalizing performance to the population. Socioeconomic level was estimated by the number of years of education completed or in the case of children, their parent's education level. Other technical characteristics, such as reliability (internal consistency, stability, and interrater agreement) and validity of SB-5 scores were generally considered positive in two independent reviews (Johnson & D′Amato, 2005; Kush, 2005). Both reviews noted improvements over the SB-4 but both also noted some problems.

INDEPENDENT RESEARCH WITH STANFORD-BINET INTELLIGENCE SCALES-FIFTH EDITION

Gale H. Roid (2003a, 2003b, 2003c) claimed the Stanford-Binet Intelligence Scales-Fifth Edition (SB-5) measured five CHC intelligence factors within verbal and nonverbal domains based on the test design and on use of confirmatory factor analysis (CFA) procedures. Factor analysis includes several approaches to investigate

how a variety of measures relate and thus, how they define underlying dimensions. Generally, CFA procedures examine competing structural models to see which fits the data best while exploratory factor analysis (EFA) examines a correlation matrix to determine how many factors or dimensions should be extracted and retained to reflect the underlying structure.

Independent studies have seriously challenged the claim that the SB-5 measures five-factors using SB-5 standardization data. Christine DiStefano and Stefan C. Dombrowski (2006) recognized the problem of not using, or reporting, EFA and set to rectify this in their study of the SB-5 standardization data. They used both EFA and CFA procedures to determine the underlying structure of the SB-5. None of the analyses (EFA or CFA) performed by DiStefano and Dombrowski on the SB-5 standardization sample data found evidence for a five-factor model and only modest support was found for two-factors (verbal and nonverbal) and only with the two youngest age groups. The verbal and nonverbal dimensions were so moderately (EFA) to highly (CFA) correlated, DiStefano and Dombrowski concluded that the SB-5 was probably best explained as a unidimensional test of intelligence.

Gary L. Canivez (2007a, 2007b), like DiStefano and Dombrowski (2006), also failed to find empirical evidence for a five-factor model and further investigated the viability of the two-factor (verbal & nonverbal) SB-5 model for the child and adolescent subsamples from the SB-5 standardization sample. Using a hierarchical exploratory factor analysis method (Schmid & Leiman, 1957), which is a recommended procedure to understand how variance is apportioned at different interpretive levels (Carroll, 1993; Carretta & Ree, 2001), Canivez (2007a, 2007b) found that the overwhelming majority of variance measured by the SB-5 was at the general intelligence factor level (Carroll's Stratum III) and little variance seems to be measured at the lower level (verbal and nonverbal, Carroll's Stratum II). Failure to find support for Roid's (2003c) five factors and limited support for even two factors (Canivez, 2007a, 2007b; DiStefano & Dombrowski, 2006) may be the result of Roid extracting too many factors due to not considering exploratory factor analyses and multiple factor extraction criteria (Frazier & Youngstrom, 2007). As a result, clinical interpretation of the SB-5 should primarily reside at the global, general intelligence level until research adequately supports interpretation of lower-order (Stratum II) dimensions.

In summary, the SB-5 appears to be a very good measure of general intelligence across a wide age range, and the standardization sample appears to be a close match to the population on key demographic variables. IQ scores also cover a wide range of ability from the lower levels of moderate mental retardation to the higher levels of intellectual giftedness. As such, they will be helpful in assessing students with mental retardation, learning disabilities, and intellectual giftedness, and interpretation of the global Full Scale IQ appears to have strong empirical support. However, as of 2007 much more research is required before interpretation of two- or five-factor models can be supported or recommended.

SEE ALSO *Intelligence Testing.*

BIBLIOGRAPHY

Binet, A. (1911). Nouvelle recherches sur la mesure du niveau intellectuel chez les enfants d'école. *L'Annee Psychologique, 17,* 145–210.

Binet, A., & Henri, V. (1895). La psychologie individuelle. *L'Annee Psychologique, 2,* 411–465.

Binet, A., & Simon, T. (1905). Methodes nouvelles pour le diagnostic du niveau intellectual des anormaux. *L'Annee Psychologique, 11,* 191–244.

Binet, A., & Simon, T. (1908). Le developpement de l'intelligence chez les enfants. *L'Annee Psychologique, 14,* 1–94.

Canivez, G. L. (2007a). Orthogonal higher-order factor structure of the Stanford-Binet Intelligence Scales for children and adolescents. Manuscript submitted for publication.

Canivez, G. L. (2007b, August). Hierarchical Factor Structure of the Stanford-Binet Intelligence Scales-Fifth Edition. Paper presented at the 2007 Annual Convention of the American Psychological Association, San Francisco, CA.

Carretta, T. R., & Ree, J. J. (2001). Pitfalls of ability research. *International Journal of Selection and Assessment, 9,* 325–335.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies*. New York: Cambridge University Press.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin, 40,* 153–193.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54,* 1–22.

Cattell, R. B., & Horn, J. L. (1978). A check on the theory of fluid and crystallized intelligence with description of new subtest designs. *Journal of Educational Measurement, 15,* 139–164.

DiStefano, C., & Dombrowski, S. C. (2006). Investigating the theoretical structure of the Stanford-Binet-Fifth Edition. *Journal of Psychoeducational Assessment, 24,* 123–136.

Frazier, T. W., & Youngstrom, E. A. (2007). Historical increase in the number of factors measured by commercial tests of cognitive ability: Are we overfactoring? *Intelligence, 35,* 169–182.

Johnson, J. A., & D′Amato, R. C. (2005). Review of the Stanford-Binet Intelligence Scales: Fifth Edition. In R. S. Spies & B. S. Plake (Eds.), *The sixteenth mental measurements yearbook* (pp. 976–979). Lincoln, NE: Buros Institute of Mental Measurements.

Kaufman, A. (1990). *Assessing adolescent and adult intelligence.* Boston: Allyn & Bacon.

Kush, J. C. (2005). Review of the *Stanford-Binet Intelligence Scales: Fifth Edition*. In R. S. Spies & B. S. Plake (Eds.), *The sixteenth mental measurements yearbook* (pp. 979–984). Lincoln, NE: Buros Institute of Mental Measurements.

Roid, G. H. (2003a). *Stanford-Binet Intelligence Scales: Fifth Edition.* Itasca, IL: Riverside.

Roid, G. H. (2003b). *Stanford-Binet Intelligence Scales: Fifth Edition, Examiner's Manual.* Itasca, IL: Riverside.

Roid, G. H. (2003c). *Stanford-Binet Intelligence Scales: Fifth Edition, Technical Manual.* Itasca, IL: Riverside.

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22*, 53–61.

Terman, L. M. (1916). The measurement of intelligence. Boston: Houghton Mifflin.

Terman, L. M., & Merrill, M. A. (1937). *Measuring intelligence.* Boston: Houghton Mifflin.

Terman, L. M., & Merrill, M. A. (1960). *Stanford-Binet Intelligence Scale.* Boston: Houghton Mifflin.

Terman, L. M., & Merrill, M. A. (1973). *Stanford-Binet Intelligence Scale: 1972 Norms Editions.* Boston: Houghton Mifflin.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Stanford-Binet Intelligence Scale: Fourth Edition.* Chicago: Riverside.

Gary L. Canivez

STEREOTYPE THREAT

According to Steele and Aronson (1995), stereotype threat is defined as a "socially premised psychological threat that arises when one is in a situation or doing something for which a negative stereotype about one's group applies" (Steele, 1997, p. 614). Another description of stereotype threat suggests that individuals are at risk of confirming a negative stereotype about their group. Here, individuals who experience stereotype threat are 1) acknowledging that a negative stereotype exists (i.e., salient in a given context or is explicitly stated) about the capabilities of their social group (i.e., race/ethnicity, gender, age, socioeconomic status) and 2) demonstrating apprehension about confirming the negative stereotype by engaging in particular activities.

An example of stereotype threat is a member of a stigmatized group (i.e., African American students, women) feeling apprehension about performing on an academic task because the individual is afraid that a possible poor performance may confirm a pre-existing negative stereotype about the individual's group (i.e., intellectual capabilities of African Americans or perceived underperformance of women in science and mathematics). For Steele, it is unnecessary for the group member to believe the stereotype to be true for stereotype threat to produce negative psychological consequences for the individual. That is, the psychological reactions to stereotype threat—exposure to contexts in which negative stereotypes about the capabilities and behaviors of a given group are or have been salient—are enough to alter the attitudes and behaviors of individual group members and produce maladaptive psychological functioning.

TASK PERFORMANCE SUBVERSION

Much of the research on stereotype threat has shown that the task performance of otherwise capable individuals is hindered when such a social-psychological threat is presented at the time of the performance (Aronson et al., 1999; Steele & Aronson, 1995; Steele, 1997). Steele (1997) and Aronson (2002) write that, for many stigmatized groups—namely women and ethnic minority populations—stereotype threat is a common reality. In particular, low-income African American and Latino students are often exposed to academic contexts in which, historically, negative beliefs regarding their perceived intelligence have been held. The awareness and salience of the belief regarding their intelligence can disrupt academic performance for these students.

The consequences of stereotype threat have been noted. For example, in a review, Aronson (2002) notes that perceptions of negative stereotypes lead many individuals to engage in activities such as self-handicapping (Smith, 2004), challenge avoidance (Good, Aronson, & Inzlicht, 2003), self-suppression (Steele, 1997; Pronin, Steele, & Ross, 2002), and disidentification or disengagement with the task or the context in which the task is to be performed (Steele, 1997; Aronson, 2002; Major et al., 1998). In addition to these poor academic performance correlates, stereotype threat has also been linked to high blood pressure among African Americans (Blascovich et al., 2001), altered career and/or professional aspirations and belonging (Steele, James, & Barnett, 2002), and social distancing, particularly from the stigmatized social group of which the participants are members (Pronin, Steele, & Ross, 2002).

These psychological and behavioral outcomes found among low-income African American and Latino and women students are not typically the result of negative stereotypes being communicated directly to them from others within the given social context. Rather, these behaviors typically result from exposure to a context in which historically 1) the performance of a given group is evaluated and compared with that of others, 2) such performance has been valued by the group and the larger society, and 3) the performance of one's group has been consistently negatively evaluated and thus stereotyped more than other groups.

MAJOR FINDINGS ON STEREOTYPE THREAT IN ACADEMIC DOMAINS

Much of the support for the presence and effects of stereotype threat has been garnered through experimentally designed studies that have been based on the