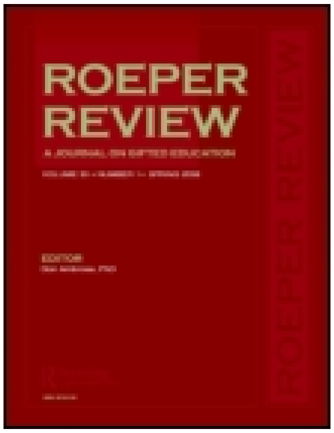
This article was downloaded by: [The Aga Khan University]
On: 21 November 2014, At: 08:14
Publisher: Routledge
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Roeper Review

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uror20

Expanding Talent Search Procedures by Including Measures of Spatial Ability: CTY's Spatial Test Battery

Heinrich Stumpf , Carol J. Mills , Linda E. Brody & Philip G. Baxley

Published online: 10 Oct 2013.

To cite this article: Heinrich Stumpf , Carol J. Mills , Linda E. Brody & Philip G. Baxley (2013) Expanding Talent Search Procedures by Including Measures of Spatial Ability: CTY's Spatial Test Battery, Roeper Review, 35:4, 254-264, DOI: 10.1080/02783193.2013.829548

To link to this article: http://dx.doi.org/10.1080/02783193.2013.829548

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the “Content”) contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions

Roeper Review, 35:254–264, 2013

ISSN: 0278-3193 print / 1940-865X online
DOI: 10.1080/02783193.2013.829548

Routledge
Taylor & Francis Group

Expanding Talent Search Procedures by Including Measures of Spatial Ability: CTY's Spatial Test Battery

Heinrich Stumpf, Carol J. Mills, Linda E. Brody, and Philip G. Baxley

The importance of spatial ability for success in a variety of domains, particularly in science, technology, engineering, and mathematics (STEM), is widely acknowledged. Yet, students with high spatial ability are rarely identified, as Talent Searches for academically talented students focus on identifying high mathematical and verbal abilities. Consequently, students with high spatial abilities who do not also have high math or verbal abilities may not qualify. In an effort to identify students with spatial talent, the Center for Talented Youth developed a Spatial Test Battery to supplement its mathematical and verbal Talent Searches. This article traces the development of the battery; describes its components, important psychometric properties, and continuing development; and encourages its use by researchers and educators interested in developing spatial talent.

Keywords: block rotation test, CTY Spatial Test Battery, spatial ability, spatial test, STEM, surface development test, talent search, visual memory test

Despite the fact that a number of spatial tests have fared well as predictors of performance in many fields (e.g., Hegarty & Waller, 2005; Lohman, Pellegrino, Alderton, & Regian, 1987), spatial ability has long been an area of cognitive functioning that is grossly neglected in educational admissions testing in general (e.g., Gohm, Humphreys, & Yao, 1998; Snow, 1999) and in talent searches to identify academically talented students in particular (e.g., Webb, Lubinski, & Benbow, 2007). The focus, instead, has been more on assessing mathematical and verbal reasoning abilities to predict achievement, even in fields such as science where requirements go beyond these two domains. In fact, it may be that the relative success of using measures of verbal and mathematical reasoning abilities, such as the SAT, to predict achievement in the educational domain has diverted attention from the testing of other abilities.

In some contexts, it can be difficult to effectively demonstrate the detrimental cost of neglecting the identification of students high in spatial ability. For this reason, the consequences are typically inferred *ex negativo*, by noting the neglect and stressing the potential benefits of the use of spatial tests. Yet, there is also direct evidence of the negative effects of not recognizing spatial skills. The German researcher Sturzebecher (1972) was among the first to contribute such observations. His data indicated that the predominantly verbal and mathematical (explicit and implicit) selection mechanisms used in schools kept many spatially talented students from reaching the university level, where they could have successfully used their talent in many science, technology, engineering, and mathematics (STEM) domains. In a similar vein, Gohm et al. (1998) argued that spatially talented students underachieved in high school and college compared to their mathematically talented peers because instructors and counselors typically employed strategies that placed an emphasis on the mathematical and verbal skills measured in college admission tests.

It is now widely believed that spatial ability may be of limited relevance for a range of highly verbal subjects and professions (such as languages, history, and journalism) but is *critically* important in many others, especially STEM fields. Support for this claim comes from four major sources (see Stumpf, 2006, pp. 15–34, for a more detailed overview):

1. A heterogeneous collection of studies arguing that many cognitive strategies used for studying STEM subjects and professional work in these fields are spatial, including the following:

Accepted 23 October 2012.

Address correspondence to Linda E. Brody, CTY, Johns Hopkins University, 5801 Smith Avenue, McAuley Hall #400, Baltimore, MD 21209. E-mail: lbrody@jhu.edu

   a. Analyses of cognitive requirements of typical scientific tasks, such as the translation of chemical formulas into three-dimensional mental models of molecules (Barke, 1995) or the derivation of models from empirical observations (Barnea, 2000).
   b. Analyses of the spatial strategies for problem solving in STEM, such as mental rotation in mathematics (Lehmann & Jüling, 2002) or the shifting between spatial and nonspatial approaches in talented students' math learning (Krause et al., 1999).
   c. Factor-analytic research on cognitive abilities relevant for the study and practice of physics (Kozhenikov, Hegarty, & Mayer, 2002; Peltzer, 1988).
2. Research on the predictive validity of spatial tests with respect to success in math and science courses (reviewed in Stumpf, 2006, pp. 15–31) and with regard to the choice of careers in the STEM fields, to acquire degrees in them, and to work in them (e.g., Lubinski, 2010; Shea, Lubinski, & Benbow, 2001; Wai, Lubinski, & Benbow, 2009). These latter studies have received much attention and have greatly reinforced interest in spatial ability.
3. The success of spatially oriented training programs for improving academic performance in a number of STEM domains, such as introductory engineering (Martín-Gutiérrez et al., 2010) and surgery (Tendick et al., 2000).
4. The observation that training in math and science can also result in improved spatial skills. That, of course, is programmatic in geometry (e.g., Kaufmann, Steinbügl, Dünser, & Glück, 2005; Maier, 1999, Chapters 7 & 8) but applies to other sciences as well, such as biology and physics (e.g., Lennon, 2000; Pallrand & Seeber, 1984).

Despite the vast amount of empirical evidence that spatial ability is important for high achievement in many STEM fields and the insistence of many experts on the potential merits of utilizing spatial tests in education (Smith, 1964, Chapter 4; see Super & Bachrach, 1957), educational administrators continue to be reluctant to include spatial measures in their assessments, especially admission tests (Snow, 1999; but see Michel et al., 1977, for a notable exception). Thus, spatial ability is still considered a "sleeping giant for talent identification and development" (Lubinski, 2010, p. 345). For the past 2 decades, however, the Center for Talented Youth (CTY) at Johns Hopkins University has been developing and launching a Spatial Test Battery (STB) to change that, at least for the purpose of identifying young adolescents for its STEM courses. The goal is to supplement CTY's Talent Searches, which use above-grade-level verbal and mathematical reasoning tests to identify students who have advanced academic abilities with an above-level test of spatial ability.

THE CTY TALENT SEARCHES

CTY has sponsored annual Talent Searches since 1979. Students who score well on grade-level standardized tests (currently at or above the 95th percentile) are invited to take aptitude tests designed for students much older than they are. Those who achieve established requirements on the above-grade-level tests are eligible to take challenging coursework offered by CTY in the humanities, writing, mathematics, science, engineering, and computer science.

This model of talent identification and development, which has its roots in the work of Julian Stanley and the Study of Mathematically Precocious Youth (SMPY), has proven to be highly successful. SMPY found that the SAT-M, administered out-of-level to middle school students, is a valid predictor of performance in advanced mathematical coursework. CTY expanded the Talent Search to include identification of high-verbal abilities and assessment of a broader age group, and the Talent Search model became the basis of other programs for academically talented precollege students in the United States and abroad (Brody, 2009; Touron, 2005).

Although CTY's Talent Searches continue to use the SAT as the primary means of identification for seventh and eighth graders, additional tests are also accepted, including the ACT for seventh and eighth graders and the School and College Ability Test (SCAT) for second through eighth graders. However, beginning in 1996, CTY began to broaden its identification model further by offering an assessment of spatial ability. Following an extensive period of development, its STB could be used in conjunction with a student's math and verbal scores on the SAT, ACT, or SCAT to establish program eligibility for CTY programs. In 2012, an initiative was undertaken to allow students with *exceptionally* high STB scores to qualify for CTY STEM classes with the STB alone. The performance of students who enrolled in courses via this option is being evaluated.

DEVELOPMENT OF THE SPATIAL TEST BATTERY

The history of the development of the STB first goes back to 1991 when, with funding from two major foundations, CTY began a multiyear initiative to explore the role that spatial ability might play in predicting high achievement in the sciences and mathematics. The basic motivation for this effort was the growing concern with regard to the competitive status of the United States in the fields of science and technology. In response to this challenge, CTY sought to expand and (as far as possible) improve its methods for identifying STEM talent.

CTY adopted a three-prong approach to investigating how to proceed that included (a) hosting a 3-day conference with leading scientists and educators who specialized in the area of scientific talent, (b) interviewing leading scientists, and (c) conducting a thorough review of the literature. The conclusion from these efforts was that it is not a simple process to identify future STEM innovators, but a central theme emerged suggesting that spatial ability plays a critical role in scientific innovation. In addition, CTY felt that, of all of the factors that were identified, spatial ability was something it could objectively measure through its Talent Searches and nurture by offering students with high-spatial ability an opportunity to engage in challenging coursework in math, science, and engineering. At this point, CTY began the process of developing a viable instrument to measure spatial ability.

To begin, it was necessary to identify those particular types of spatial tests that were likely to serve CTY's objectives best. The literature review found studies that came from many different STEM domains and had used many different types of spatial tests and criteria of educational success. Thus, we began by incorporating many of these types into our initial attempts at developing an STB.

The first iteration of the STB was composed of two forms. Each form had 14 subtests, which represented 14 different types of spatial tasks in terms of Eliot's (1980, 1983) classification of figural spatial tests. Most of the 28 subtests in these two initial STB forms had been published and used before. Each of the two STB forms was administered to a separate sample of CTY students who were taking courses in mathematics or the sciences. Teacher ratings of performance in the courses were obtained. Most of the 28 subtests proved to have adequate internal consistency. A strong general factor was found to underlie performance on the two forms (Stumpf & Eliot, 1995). Similarity structure analyses (SSAs; Borg & Lingoes, 1987) yielded a RADEX-like structure with complex, three-dimensional types of tasks defining the core of spatial ability and two-dimensional, mostly speed-oriented, tasks represented in the periphery of the solutions (Stumpf & Eliot, 1999). The tasks in the core of the SSA representations also tended to have the higher predictive validities with respect to course performance. Several types of tests, namely, Surface Development, Paper Folding, Block Rotation, Intersections, and Perspectives tasks, showed a convergence of good psychometric properties: They yielded encouraging instances of predictive validity, demonstrated at least adequate internal consistency (with alpha coefficients ranging from .73 to .96), had high loadings on the general spatial factor, and were represented in the core of the SSA solutions.

Tests of the above types were retained for inclusion in several new, shorter forms of the STB (see Stumpf, 2006, pp. 35–38, for an overview), and new items were created by CTY researchers. Research with the short forms demonstrated that selected subtests were appropriate for students below the seventh-grade level. Among the new versions of the STB was Form HH, the first STB version that consisted entirely of items newly constructed at CTY. In a study with 423 sixth graders at international schools (Stumpf & Haldimann, 1997), Cronbach alpha coefficients for these tests ranged from .75 to .89, with an alpha of .90 for the total score. The latter score was found to have a correlation of .39 with the GPA. The forms of the STB that were created up to this time were designed to be completed in a paper-and-pencil format and administered by proctors. However, for practical use in a Talent Search, a way had to be found to offer the test more widely and in communities where students lived. In 1994–1995, an opportunity to computerize the test presented itself, which would allow students to take the test at a test center near their home.

COMPUTERIZATION OF THE STB

In constructing the new computerized version of the STB, the plan was to pilot it as a new diagnostic component of CTY's Talent Searches for its seventh-grade and above STEM courses. The new version of the STB had to be confidential, short enough to be administered in one session and in a reasonable amount of time, and appropriate for computerized administration. The confidentiality requirement necessitated that all items for the planned STB version had to be newly developed but in a way that they represented the item types that had been examined before. The time requirement limited the number of subtests to about four.

Among those item types that had shown encouraging properties were Paper Folding and Surface Development, but both require the imagination of folding and unfolding. So, to avoid redundancy, only one type could be selected from this area; we chose Surface Development. In view of the key role of mental rotation in spatial ability, it was imperative to keep Block Rotation. From the remaining types, Perspectives and Intersections, we decided to include only the first. Intersections might also have been a good choice, but it was left out to conserve testing time. Thus, the decision was to newly construct subtests that included Block Rotation, Surface Development, and Perspectives tasks. These types represent the broad spatial Visualization Factor as described by Carroll (1993, Chapter 8), which encompasses the components Spatial Visualization and Spatial Orientation, which had been identified in previous research as separate factors.

The Visualization Factor is by far the most important spatial component, but it does not exhaust the whole spatial domain. We therefore decided to consider at least one other factor that has been found to be important in the context of spatial ability: Visual Memory (Ekstrom, French, & Harman, 1976; Hegarty & Waller, 2005; Lohman et al., 1987; described by Carroll, 1993, ch. 7, in the context of memory factors). In view of its practical importance, a subtest of Visual Memory was included in the new computerized versions of the STB. Thus, the computerized STB constructed in 1994–1995 consisted of four subtests: Surface

Development (SD), Block Rotation (BR), Visual Memory (VM), and Perspectives (PE).

Significant changes to the computerized STB were made on three subsequent occasions. In 1997, new designs for the Block Rotation and Perspectives subtests were introduced (without changing the type of tasks in terms of Eliot's 1980 and 1983 classifications). In 2001, an STB version for younger students (fifth and sixth graders) was introduced. Adjustments were made in 2006 to increase the number of items in the Visual Memory subtest and to reduce the number of items and time allotted for Surface Development in order to improve the psychometric properties of the test.

There are currently four forms on two difficulty levels of the computerized STB: Forms 11 and 12 (Level I) for young students (fifth and sixth grade) and Forms 21 and 22 (Level II) for older students (seventh grade and above). Forms 11 and 12 contain three subtests: Visual Memory, Surface Development, and Block Rotation. Forms 21 and 22 contain these three subtests, as well as Perspectives. All forms contain counting items and experimental items, which can be evaluated for future use. Experimental items are used to *refresh* counting items for the purpose of (a) adjusting or improving the psychometric properties of items/subtests, as needed, and (b) maintaining test security. In Forms 11 and 12, and in Forms 21 and 22, the counting items are identical. On each difficulty level, the two forms differ, however, with respect to the experimental items. CTY uses two forms at each difficulty level so that more experimental items can be tested and calibrated each year. Students are randomly assigned to a form within the appropriate difficulty level. Table 1 shows the number of minutes allowed and the number of questions for each subsection.

TABLE 1
Numbers of Items and Time Allowances for the STB Subtests

	Level I		*Level II*	
	Minutes	*Items*	*Minutes*	*Items*
Visual Memory Memorization Phase: Shape Viewing	8	15 Shapes	8	22 Shapes
Surface Development subtest	12	30 (6 pairs of shapes with 5 questions per pair)	12	30 (6 pairs of shapes with 5 questions per pair)
Block Rotation subtest	12	20	9	20
Visual Memory Recall Phase: subtest	8	15	8	20
Perspectives subtest	N/A	N/A	19	22
Totals	83	65	109	92

Note. The total test times include the time allotted for the actual subtests, the tutorials for the whole battery and the subtests, and the questionnaires presented at the beginning and end of the testing session. The latter are not indicated in this table. A complete list can be found at http://cty.jhu.edu/talent/testing/about/stb.html (Johns Hopkins Center for Talented Youth, 2013). Fifth- and sixth-grade students take Level I; seventh-grade and above students take Level II.

Before the actual test battery is administered, a general tutorial explains how to use the computer and familiarizes the student with the overall layout of the test. This is followed by a short sociodemographic questionnaire and detailed subsection tutorials that include moving images preceding the Surface Development, Block Rotation, and Perspectives subtests. The Visual Memory subtest has two parts, a memorization phase presented at the beginning of the actual battery, and a recognition section presented after the Block Rotation subtest. Both parts are preceded by conventional instructions. The tutorials, instructions, and sample items are available on the Internet at http://cty.jhu.edu/talent/testing/about/stb.html (Johns Hopkins Center for Talented Youth, 2013).

Within each subtest, easy tasks are displayed first and more difficult tasks later. All items are multiple-choice, and each item has only one correct solution. Students can skip questions within a subtest and go back to change an answer, but once a subtest is exited it cannot be reentered. Each subtest is timed separately, and a timer, located in the lower left-hand portion of the screen, displays the time remaining as students work through the questions. The experimental items, which do not count toward the student's score and are included only to be calibrated for future use, are not recognizable by the students as such. If these items have acceptable difficulty and discrimination indices, they are eligible to replace previously counting items in the next testing cycle.

PSYCHOMETRIC PROPERTIES OF THE STB

Difficulty, Internal Reliability, and Item Discrimination

Table 2 summarizes essential psychometric properties of the STB subtests and the total STB for Levels I and II, means and standard deviations of the raw scores, average item p-values, and Cronbach alpha coefficients of internal consistency. Also displayed in Table 2 are the average item discrimination indices for both levels of the STB. The statistics refer only to the counting items, not the experimental ones. Forms 11 and 12 are combined, as are Forms 21 and 22. The indices presented are based on data from four groups of students:

- Group 1: $N = 693$ (276 females, 417 males), fifth and sixth graders, 2004 Talent Search
- Group 2: $N = 1004$ (376 females, 728 males), seventh and eighth graders, 2004 Talent Search
- Group 3: $N = 299$ (103 females, 196 males), fifth and sixth graders, 2011 Talent Search
- Group 4: $N = 320$ (108 females, 212 males), seventh and eighth graders, 2011 Talent Search

The average p-values for the various subtests displayed in the table are near the optimal value of .5. The only exception is the Surface Development subtest on Level II in the

TABLE 2
Psychometric Properties of the STB Subtests and the STB Composite

						Average Item Discrimination Index		
		Mean	*SD*	*Avg.* p	*Alpha*	*Biserial*	*Point Biserial*	*Point Biserial*
Level I (Forms 11 and 12)								
Surface Development	Group 1	13.62	6.27	.54	0.90	0.73	0.55	0.49
	Group 3	13.95	6.71	.56	0.91	0.74	0.57	0.52
Block Rotation	Group 1	9.54	4.27	.53	0.82	0.64	0.50	0.40
	Group 3	10.62	4.39	.59	0.83	0.66	0.51	0.42
Visual Memory	Group 1	6.62	3.72	.44	0.79	0.64	0.51	0.40
	Group 3	8.61	3.70	.57	0.81	0.67	0.52	0.42
Total STB	Group 1	29.78	11.41	.51	0.92	0.54	0.42	0.38
	Group 3	33.18	11.95	.57	0.92	0.56	0.43	0.40
Level II (Forms 21 and 22)								
Surface Development	Group 2	16.01	6.43	.64	0.91	0.73	0.56	0.50
	Group 4	11.40	5.58	.57	0.90	0.77	0.59	0.53
Block Rotation	Group 2	10.63	3.65	.59	0.77	0.61	0.45	0.34
	Group 4	10.23	4.28	.57	0.83	0.66	0.50	0.42
Visual Memory	Group 2	9.63	4.85	.48	0.83	0.62	0.49	0.41
	Group 4	10.32	4.43	.57	0.83	0.65	0.51	0.42
Perspectives	Group 2	9.91	3.98	.50	0.74	0.53	0.41	0.30
	Group 4	10.95	4.51	.55	0.81	0.55	0.46	0.37
Total STB	Group 2	46.18	14.09	.56	0.92	0.47	0.36	0.33
	Group 4	42.90	14.43	.56	0.93	0.51	0.40	0.37

Note. Two point-biserial discrimination indices are specified. The second index is part–whole corrected; the first is not corrected for overlap.

2004 Talent Search (Group 3), which was fairly easy. This was corrected by replacing easy counting items with more difficult experimental items. With alpha coefficients ranging from .74 to .91, the internal consistency of all subtests is also appropriate according to the common standards (e.g., Nunnally, 1978) and, in many cases, high. The total scores on each difficulty level possess very high degrees of internal reliability (alphas ranging from .92 to .93). The good internal consistency of the subtests is also highlighted by the average item discrimination indices shown in the table. On the item level, difficulty and discrimination indices are monitored as old items are replaced by new ones. Overall, the process of item substitution preserves the internal consistency of the subtests and allows CTY the option of adjusting the difficulty level of subtests, as appropriate. A slight secular trend toward improvement in test performance has been noted. This trend is being monitored and, if needed, corrections will be made in the next revision.

Correlations With the SCAT and the SAT

Table 3 shows the correlations of the standard scores on the STB subtests and the total STB with the standard scores on the SCAT and the SAT. Also included in the table are the correlations of the STB subtests with each other. The statistics are based on data from participants in the CTY Talent Searches of 2008 through 2011.

TABLE 3
Intercorrelations of the Standard Scores on the STB Subtests, the Total STB, and the Scores on the SCAT and the STB

		(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*
STB, Level I:									
Surface Development	(1)	1.00							
Block Rotation	(2)	0.56	1.00						
Visual Memory	(3)	0.36	0.31	1.00					
STB-SSC	(4)	0.87	0.79	0.63	1.00				
SCAT Math	(5)	0.41	0.37	0.25	0.46	1.00			
SCAT Verbal	(6)	0.33	0.31	0.24	0.38	0.54	1.00		
SCAT Total	(7)	0.43	0.39	0.29	0.48	0.91	0.85	1.00	
STB, Level II:									
Surface Development	(1)	1.00							
Block Rotation	(2)	0.57	1.00						
Visual Memory	(3)	0.29	0.19	1.00					
Perspectives	(4)	0.54	0.45	0.30	1.00				
STB-SSC	(5)	0.84	0.73	0.59	0.76	1.00			
SAT-I M	(6)	0.45	0.34	0.24	0.42	0.51	1.00		
SAT-I CR	(7)	0.25	0.18	0.21	0.30	0.32	0.50	1.00	
SAT-I Total	(8)	0.41	0.30	0.26	0.42	0.48	0.88	0.86	1.00

The results presented for Level I in Table 3 are based on data from 3,067 students who had completed both the SCAT and Level I of the STB (Forms 11 or 12) in the Talent Searches mentioned in the previous paragraph. Of these students, 51.3% were fifth graders and 48.7% were sixth graders; 35.9% were female. If multiple test scores were available for a student, only the STB scores obtained in his or her first testing were included; the SCAT scores used were those obtained in the testing that was closest in time to the STB testing.

The results for Level II are based on the performance of 1,554 students who took both the SAT and Level II (Form 21 or 22) of the STB. Ninety-eight percent of the students in this group were seventh or eighth graders; 36.2% were female. The procedure for including participants having multiple STB and/or SAT scores was analogous to that described in the previous paragraph.

As far as the STB subtests are concerned, the correlations among the Surface Development, the Block Rotation, and, on Level II, the Perspectives subtest, were found to be the highest, with Visual Memory showing markedly lower correlations with the other parts of the STB than the three subtests mentioned first among each other. As markers of the broad Visualization Factor described by Carroll (1993), those three tests have more in common with each other than with the Visual Memory subtest.

The correlations of the STB scores with the scores on the SCAT and the SAT reiterate a frequently reported observation: Spatial tests tend to show stronger covariation with mathematical tests than with verbal tests. Here, however, the correlations are in the upper range of what one would expect from previous meta-analyses (Friedman, 1995), and they are not necessarily lower than those between verbal and mathematical reasoning tests. The pattern of correlations shown in the tables is familiar from previous studies with different versions of the STB, the SAT, and other verbal and mathematical reasoning tests (e.g., Stumpf, 1998, 2006; Stumpf & Eliot, 1995; Stumpf & Haldimann, 1997).

Predictive Validity

From the outset, it was clear that CTY's use of existing tests of verbal and mathematical reasoning as the primary means for student identification meant that a large part of the predictive variance had already been captured. In addition, most of the criterion measures available to evaluate course success focus on verbal and mathematical tasks. Like most efforts to extend knowledge in the STEM domains, the CTY STEM courses involve the teaching of factual information and multiple skills to solve problems in the various STEM fields. Some of these skills are spatial (such as technical drawing in engineering and observation techniques in the microscopy courses), others are nonspatial (such as techniques for efficiently acquiring and organizing new factual information and using logical reasoning for generating new insights). The acquisition of knowledge and the development of both types of skills should be the *criterion* of a student's performance in a course. Unfortunately, the criterion *measures* used by CTY instructors so far mainly focus on nonspatial criteria at the expense of the spatial skills developed. Rigorous tests of the latter criteria still need to be introduced. For this reason, the available data on the predictive validity of the STB can be regarded only as preliminary and incomplete.

With this situation and understanding in mind, two predictive validity studies with the STB have been completed as of now to address the question of whether the STB adds incremental validity to the prediction made by the other tests used in the Talent Searches. The studies were based on data of students who had been admitted to CTY mathematics and science courses through the regular Talent Search process. In both studies, the criterion measure of study success was a 15-item rating form used by CTY to assess course performance of its students. Every CTY teacher completed the 15 Likert-type ratings (e.g., "Understood complex material," "Solved problems well") for every student in her or his class at the end of each course. The rating form was used in all courses (including the humanities) and thus did not specifically address qualifications in particular sciences or spatial skills. A strong general factor was found to underlie the 15 ratings, accounting for about 60% of the variance. This had two important consequences for the validity studies: On the one hand, the total score on the rating form (average across the ratings made by the teacher, provided that at least 14 items were completed) had a very high degree of internal consistency (with alpha coefficients above .90 in both studies). Thus, the criterion measure was highly reliable. On the other hand, the score reflected only a very general evaluation of student performance, without addressing specific strengths or weaknesses. Another problem of the criterion measure was that the distributions of the scores were highly skewed: More than 80% of the students completing a STEM course attained average ratings in the two highest categories (4 and 5). Thus, given the fact that the distributions of the predictor scores—the STB, SAT, and PLUS[1] tests—deviated much less from normality, the shapes of the score distribution alone substantially limited the potential correlations of all predictors with the criterion measure.

The first study was conducted in 1998 with seventh-grade and above CTY students. A paper–pencil version of the computerized STB was administered to incoming CTY students at various summer program sites and the most recent SAT scores of all students at those sites were obtained. Table 4 displays the predictive validity coefficients of the STB and the SAT in this CTY population separately for mathematics and computer science students, for science students, and for the combined math/science group. Table 5 shows results

TABLE 4
Correlations of the Predictor Scores With the Average Rating in the First Predictive Validity Study

	Mathematics/Computer Science	*Science*	*Total Group*
STB Total	0.41 (67)	0.21 (113)	0.29 (180)
SAT-I Math	0.13 (603)	0.05 (743)	0.08 (1,421)
SAT-I Verbal	0.15 (604)	0.16 (743)	0.16 (1,422)
SAT-I Total	0.16 (603)	0.13 (743)	0.14 (1,421)

Note. The number of cases on which the correlations are based is given in parentheses.

TABLE 5
Results of the Regression Analysis With the SAT-I and the STB as Predictors of Performance in the CTY Mathematics and Science Courses

	Multiple Correlations	*Beta Weights*	
Mathematics and Computer Science ($N = 66$)			
SAT-M	0.33	SAT-M	0.10
SAT-M, SAT-V	0.36	SAT-V	0.08
SAT-M, SAT-V, STB	0.46	STB	0.22
Science ($N = 110$)			
SAT-M	0.14	SAT-M	0.17
SAT-M, SAT-V	0.14	SAT-V	0.14
SAT-M, SAT-V, STB	0.22	STB	0.31
Total Group ($N = 176$)			
SAT-M	0.22	SAT-M	0.05
SAT-M, SAT-V	0.24	SAT-V	0.00
SAT-M, SAT-V, STB	0.31	STB	0.19

from step-wise multiple regression analyses with SAT-M alone entered, with SAT-V added, and with the STB score added. The rightmost column of this table contains beta weights from simultaneous multiple regression analyses with the STB and the two parts of the SAT as predictors.

The predictive validity coefficients for the STB with respect to the criterion score were encouraging, in particular for the math/computer science and the combined group. The fact that the SAT scores, in particular SAT-M, had lower correlations with the criterion than the STB might appear surprising at first glance. It should be noted, however, that the students had been selected into the courses based on their SAT scores; thus, there was restriction of range on these scores, especially on SAT-M. In addition, the (explicit, by selection) restriction of variance on SAT-M introduced an (implicit) restriction on the STB because of the correlation between SAT-M and STB. Still, with the STB score added to the regression models, the multiple *r*s increased. Thus, there is good reason to conclude from the multiple correlations and beta weights in Table 5 that, in this sample, the STB contributed incremental validity to what prediction was achieved by the SAT alone.

The second predictive validity study was designed to replicate the results of the first with fifth and sixth graders. It was based on data from students who completed CTY STEM courses in 2003, 2004, or 2005. In this study, the STB scores on Level I (Forms 11 and 12) were used, as were the scores on the Math and Verbal parts of the PLUS test, which was used during this time period to select students for the CTY programs. To achieve comparability of the scores from different Talent Search years, standardized scores were used for both tests. Given the fact that the students had been selected into the courses primarily on their PLUS-Math scores, substantial restriction of variance was expected, explicitly on PLUS-Math and, due to the correlation of the STB with that variable, implicitly also on the STB. Therefore, beyond computing the (restricted) raw correlations between the predictors and the criterion measure, we also obtained estimates of the unrestricted predictive validity coefficients (i.e., of the correlations that would have occurred if there had been no restriction of variance). To compute those estimates, we used Thorndike's (1947) formula for one-sided correction for restriction of range on the predictors.

TABLE 6
Correlations of the Standard Scores on the PLUS Test and the STB With the Average Performance Ratings in the Mathematics and Science Courses

	Mathematics Group	*Science Group*	*Total Sample*
PLUS Math	0.24 (0.31)	0.16 (0.20)	0.20 (0.25)
	552	550	1052
PLUS Verbal	0.21 (0.22)	0.19 (0.20)	0.20 (0.21)
	552	550	1052
STB Standard Score	0.25 (0.25)	0.14 (0.15)	0.19 (0.20)
	127	135	245

Note. The coefficients given in parentheses are corrected for restriction of range. The *N*s on which the correlations are based are shown below the respective coefficients. The *N*s for the total group are not the sums of *N*s for the Math and Science groups, because students who took both a math and a science course were counted only once in the total group.

The uncorrected and corrected predictor–criterion correlations are summarized in Table 6 separately for the students who took mathematics courses and for those who took science courses and also for the total group of study participants. Comparing the raw and the corrected correlations (the latter are given in parentheses), the reader will see that noteworthy correction effects occurred mainly for the Math part of the PLUS test.

As we found with the older students, performance in the mathematics courses was easier to predict than success in the more heterogeneous science programs. In the Math group (and the total sample), predictive validity coefficients for the PLUS-Math and the STB were moderate. In the Science group, they were fairly low but still positive. The corrected validity coefficients show that the STB is second to the mathematical reasoning test (PLUS-M) in predicting performance (see Table 6). Surprisingly, however, the difference in validity to the PLUS-M was fairly small.

As in the first study, we assessed the amount of incremental predictive validity introduced by the STB in step-wise regression analyses with the PLUS test and STB scores as predictors and the average ratings as the criterion measure. However, the additional amount of criterion variance explained by the STB beyond the PLUS test turned out to be so small that these analyses are not reported in detail here. To assess the relative contributions of the predictors in explaining the variance of the criterion score, we performed simultaneous regression analyses. The beta weights and multiple correlations of these regression models are summarized in Table 7.

TABLE 7
Results of the Simultaneous Multiple Regression Analyses With the Scores on the PLUS Test and the STB as Predictors and the Average Rating as Criterion Score (Second Predictive Validity Study)

Predictors	*Beta Weights*	*Multiple* R
Mathematics Group		
PLUS Math	0.21	0.31
STB	0.16	
Science Group		
PLUS Math	0.10	
PLUS Verbal	0.10	0.22
STB	0.12	
Total Group		
PLUS Math	0.15	
PLUS Verbal	0.02	0.24
STB	0.14	

In the Math group, the weight for PLUS-Verbal turned out to be negative; therefore, that variable was dropped from the analysis for that subsample. As expected, PLUS-Math made a somewhat larger independent contribution to the prediction than the STB in the Math group, but in the Science group and the total sample, the beta weights for the two predictors were essentially the same. Thus, both the STB and the Mathematical part of the PLUS test, as single predictors, achieved a limited prediction of the criterion score, but little improvement of that prediction could be made when one predictor was added to the other.

Gender Differences

Gender-related differences in scores on the STB have been studied extensively (Stumpf, 1998; Stumpf & Eliot, 1995; Stumpf & Haldimann, 1997). Table 8 gives an overview of such differences on the most recent computerized STB versions, administered in the 2011 Talent Search. The table includes effect size indices *d*, which express the differences between the mean scores in standard deviation units. Positive *d*s indicate advantages for males here. The mean differences on Surface Development, Block Rotation, Perspectives, and the STB total score favor males, whereas those on Visual Memory favor females. According to the common criteria for the interpretation of *d* coefficients in gender differences research (e.g., Willingham & Cole, 1997), the differences on Perspectives and Block Rotation on Level II would be classified as “small to medium” and “medium,” respectively. Those on Surface Development and Block Rotation on Level I and on the two total scores would be considered “small” (but not negligible). The other mean differences are not systematic in terms of that classification.

These results are in line with findings on a paper–pencil version of the STB in an international school population (Stumpf & Haldimann, 1997). They are consistent with the general observation that, in the spatial domain, the largest gender differences occur on mental rotation tasks (e.g., Voyer, Voyer, & Bryden, 1995). However, there are advantages for females on various episodic memory tests, including certain visual tasks (e.g., Herlitz, Larsson, & Rehnman, 2004).

Differences in score means, though, are only one aspect of differential test performance. Reviews by Feingold (1992) and Willingham, Cole, Lewis, and Leung (1997) have reiterated the possibility that there are also differences in variability on many ability, aptitude, and achievement tests, with larger variances for males than for females. Such differences are not only of theoretical importance; they also have obvious practical implications whenever applicants are selected from the upper part of a score distribution. To examine that issue with the STB, we computed variance ratios. A variance ratio (as defined here) is obtained by comparing the score variances for females and males, dividing the larger variance by the smaller, and presenting the ratio in conjunction with the label of the group having the greater variability. As an example, the variance ratios shown in Table 8 for the two Visual Memory subtests indicate that females have a somewhat larger score variability on these measures. Most of the variance ratios in the table deviate little from unity; three

TABLE 8
Gender-Related Differences in Scores on the STB Subtests and the Total STB, Talent Search 2011

	Mean	SD	*Mean*	SD	d	*Variance Ratio*
Level I	Males (*N* = 196)		Females (*N* = 103)			
Surface Development	14.70	6.81	12.51	6.32	0.33	1.16 M
Block Rotation	11.06	4.43	9.80	4.20	0.29	1.11 M
Visual Memory	8.47	3.64	8.87	3.82	−0.11	1.10 F
STB Total	34.23	12.09	31.18	11.45	0.26	1.11 M
Level II	Males (*N* = 212)		Females (*N* = 108)			
Surface Development	11.58	5.51	11.05	5.73	0.09	1.08 F
Block Rotation	11.12	4.21	8.48	3.87	0.65	1.18 M
Visual Memory	10.10	4.36	10.76	4.55	−0.15	1.09 F
Perspectives	11.46	4.54	9.44	4.30	0.46	1.11 M
STB Total	44.25	14.33	40.23	14.31	0.28	1.00 M

Note. Level I contains Forms 11 and 12. Level II contains Forms 21 and 22.

of them indicate slightly larger score variability for females. Consequently, the variance ratios do not offer consistent support for the so-called greater male variability hypothesis. Thus, gender differences in variance, especially on the STB total score, are less of a concern here than that hypothesis would suggest.

DISCUSSION AND FUTURE DIRECTIONS

The importance of spatial ability for academic and career success is growing steadily as more and more fields today have strong visual–spatial components. From the air traffic controller looking at screens showing an overcrowded sky, to the scientist deciphering the mysteries of the Human Genome, to the engineer designing a bridge that will connect distant lands, excellent visual–spatial skills play a crucial role. As our world becomes even more technologically oriented in the future, it is likely that the ability to visualize and manipulate two- and three-dimensional figures will become even more important, and we need to be able to identify students with the potential to excel in this area. CTY's STB was developed for this purpose.

The STB was the result of much research and experimentation and has proven useful as part of CTY's strategy to identify students with potential to succeed in advanced and rigorous STEM courses. The battery has shown appropriate levels of test difficulty, good internal reliability, and encouraging indications of predictive validity. Gender differences in test scores are consistent with previous research. The procedure for creating updated versions of the STB from one Talent Search to the next has turned out to be very successful. Future studies on the validity of the STB will need to use improved criterion measures addressing spatial competencies and focus on more narrowly defined groups of students (such as participants in courses that reflect specific STEM domains).

The comments made by students show that the STB is well received by the large majority of the test takers. Among the features most often noted by the students is the difference in content between the STB and the more traditional educational tests that are often administered. Still, the number of students who have elected to take it each year has been much smaller than for the SAT or SCAT. This is because, prior to 2012, the STB was used only as a supplemental option in combination with the SAT or SCAT for CTY program eligibility. Now that the use of the STB as a primary qualifying test for select STEM courses is being offered, the number of students taking the test has increased and this trend is likely to continue. As the STB is used more widely by students seeking eligibility for academic programs, CTY plans to construct new courses that will utilize more fully spatial reasoning and teaching approaches.

The use of the STB has proven beneficial for Talent Search participants in several regards. For students whose math or verbal scores fail to qualify them for CTY programs, the STB option expands access to challenging math and science courses to spatially talented students who would not otherwise have been admitted. The eventual result is likely to be an increase in the talent pool identified for high achievement and innovation in STEM fields. For those who have high math and/or verbal scores and do not need the STB for program eligibility, the additional knowledge provided by understanding their specific spatial abilities can guide college and career decision making.

Though only the computerized version of the STB is used for its program eligibility, CTY licenses one remaining paper-and-pencil form of the STB, Form C, to individuals and groups for research purposes and for use with local populations. Such populations include mathematically talented students in Hong Kong and middle-school students in Egypt, Kazakhstan, Spain, Thailand, and Turkey. Form C, which contains the Visual Memory, Surface Development, and Block Rotation subtests, has been translated into Turkish, Thai, Russian, Kazakh, German, Spanish, and Catalan.

International interest in the STB continues to increase. The STB is currently being used in a number of countries as part of their protocol to identify highly able (gifted) students. In 2011, for example, Form C was given to 2,883 students in Kazakhstan as part of a pilot program to explore whether it would be a useful addition to a battery of achievement tests used for admissions testing into special regional schools focusing on advanced instruction in mathematics and the sciences. Based on the results of this pilot, the STB was included as part of the selection process when over 10,000 students from throughout Kazakhstan took the test in 2013. The STB is also included, along with an above-level version of the SCAT, in the selection process used to identify gifted children in Egypt.

We are eager to expand the research base related to the efficacy of the STB, to use it to enhance our understanding of the role of spatial abilities in academic achievement, and to have the test more widely utilized by, and for, other populations. To make the test more widely available, work is currently underway to construct a version of the STB that will be either web-based or administered locally through the use of portable equipment with the STB software stored on a disk. This will allow the STB to be administered in schools or other sites (including international locations) where it is inconvenient or impossible for students to access a test center.

Because the primary interest in developing the STB was the identification of new talent for the CTY programs, the focus has been on discovering students who attain high scores on the battery. However, concern about the importance of spatial ability, particularly in STEM fields, has stimulated an interest in *developing* spatial abilities. A promising approach for the STB to assist in this effort would be to design adaptive score reports.

The best educational opportunities for students, whether in or out of school, are personalized (i.e., they recognize, and are designed to meet, students' individual learning needs).

But for this to occur, a clear understanding of their specific cognitive abilities, of their unique pattern of strengths and weaknesses, is needed. Talent Search testing has been extremely helpful to many students in providing this information with regard to their mathematical and verbal reasoning abilities and especially by providing an *above-level* assessment that raises the assessment ceiling for gifted students who excel at grade level. But there are cognitive abilities beyond math and verbal reasoning that are crucial to achievement that should be considered and evaluated. With the increasing emphasis on preparing students for careers in STEM fields, the most important of these is likely to be spatial ability, and students participating in Talent Searches have much to gain by augmenting their math and verbal assessments with the STB.

Meanwhile, researchers and educators still have much to learn about the role that spatial ability plays in achievement in many domains, how it interacts with other abilities, and the degree to which it can be developed through practice and training.

NOTE

1. The PLUS test was developed by ETS for CTY's fifth- and sixth-grade Talent Search and was used for several years. It has now been replaced by the SCAT.

REFERENCES

Barke, H. (1995). Chemische Strukturen und räumliches Denken [Chemical structures and spatial thinking]. *Chemie in der Schule*, *42*, 402–409.

Barnea, N. (2000). Teaching and learning about chemistry and modelling with a computer managed modelling system. In J. K. Gilbert & C. J. Boulter (Eds.), *Developing models in science education* (pp. 307–323). Dordrecht, The Netherlands: Kluwer.

Borg, I., & Lingoes, J. (1987). *Multidimensional similarity structure analysis*. New York, NY: Springer Verlag.

Brody, L. E. (2009). The Johns Hopkins talent search model for identifying and developing exceptional mathematical and verbal abilities. In L. V. Shavinina (Ed.), *International handbook on giftedness* (pp. 999–1016). New York, NY: Springer.

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

Ekstrom, R. B., French, J. W., & Harman, H. H. (1976). *Manual for kit of factor-referenced tests*. Princeton, NJ: Educational Testing Service.

Eliot, J. (1980). Classification of figural spatial tests. *Perceptual and Motor Skills*, *51*, 847–851.

Eliot, J. (1983). The classification of spatial tests. In J. Eliot & I. M. Smith, *An international directory of spatial tests* (pp. 11–15). Windsor, England: NFER-Nelson.

Feingold, A. (1992). Sex differences in variability in intellectual abilities: A new look at an old controversy. *Review of Educational Research*, *62*, 61–84.

Friedman, L. (1995). The space factor in mathematics: Gender differences. *Review of Educational Research*, *65*, 22–50.

Gohm, C. L., Humphreys, L. G., & Yao, G. (1998). Underachievement among spatially gifted students. *American Educational Research Journal*, *36*, 515–531.

Hegarty, M., & Waller, D. A. (2005). Individual differences in spatial ability. In P. Shah & A. Miyake (Eds.), *The Cambridge handbook of visuospatial thinking* (pp. 121–169). New York, NY: Cambridge University Press.

Herlitz, A., Larsson, M., & Rehnman, J. (2004, August). *Sex differences in episodic memory: A meta-analysis*. Poster session presented at the 28th International Congress of Psychology, Beijing, China.

Johns Hopkins Center for Talented Youth. (2013). *CTY's Spatial Test Battery.* Retrieved from http://cty.jhu.edu/talent/testing/about/stb.html

Kaufmann, H., Steinbügl, K., Dünser, A., & Glück, J. (2005). General training of spatial abilities by geometry education in augmented reality. *Annual Review of Cybertherapy and Telemedecine*, *3*, 65–76.

Kozhenikov, M., Hegarty, M., & Mayer, R. E. (2002). Visual/spatial abilities in problem solving in physics. In M. Anderson, B. Meyer, & P. Olivier (Eds.), *Diagrammatic representation and reasoning* (pp. 155–173). New York, NY: Springer-Verlag.

Krause, W., Seidel, G., Heinrich, F., Sommerfeld, E., Ptucha, J., Schack, B., & Gundlach, W. (1999). Multimodale Repräsentation als Basiskomponente kreativen Denkens [Multimodal representation as a basic component of creative thinking]. *Jenaer Schriften zur Mathematik und Informatik*, *29*, 129–142.

Lehmann, W., & Jüling, I. (2002). Raumvorstellung und mathematische fähigkeiten—Unabhängige konstrukte oder zwei seiten einer medaille? [Spatial ability and mathematical abilities: Independent constructs or two sides of one medal]?. *Psychologie in Erzeihung und Unterricht*, *49*, 31–33.

Lennon, P. A. (2000). Improving students' flexibility of closure while presenting biology content. *American Biology Teacher*, *62*, 177–180.

Lohman, D. F., Pellegrino, J. W., Alderton, D. L., & Regian, J. W. (1987). Dimensions and components of individual differences in spatial abilities. In S. H. Irvine & S. E. Newstead (Eds.), *Intelligence and cognition: Contemporary frames of reference* (pp. 253–312). Dordrecht, The Netherlands: Martinus Nijhoff.

Lubinski, D. (2010). Spatial ability and STEM: A sleeping giant for talent identification and development. *Personality and Individual Differences*, *49*, 344–351.

Maier, P. H. (1999). *Räumliches Vorstellungsvermögen* [Spatial ability]. Donauwörth, Germany: Auer.

Martín-Gutiérrez, J., Saorín, J. L., Contero, M., Alcañiz, M., Pérez-López, D. C., & Ortega, M. (2010). Design and validation of an augmented reality for spatial abilities development in engineering students. *Computer & Graphics*, *34*, 77–91.

Michel, L., Jäger, R., Schweizer, H., Geisel, B., Habersang, U., Kierdorf, B., . . ., Melter, A. (1977). *Hochschuleingangestest für das Studienfeld Medizin* [University entrance test for medical studies]. Bonn, Germany: Bundesminister für Bildung und Wissenschaft.

Nunnally, N. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw Hill.

Pallrand, G. J., & Seeber, F. (1984). Spatial ability and achievement in introductory physics. *Journal of Research in Science Teaching*, *25*, 507–516.

Peltzer, A. (1988). The intellectual factors believed by physicists to be most important to physics studies. *Journal of Research in Science Teaching*, *25*, 229–240.

Shea, D. L., Lubinski, D., & Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology*, *93*, 604–614.

Smith, I. M. (1964). *Spatial ability: Its educational and social significance*. London, England: University of London Press.

Snow, R. E. (1999). Commentary: Expanding the breadth and depth of admissions testing. In S. Messick (Ed.), *Assessment in higher education* (pp. 133–140). Hillsdale, NJ: Lawrence Erlbaum.

Stumpf, H. (1998). Gender-related differences in academically talented students' scores and use of time on tests of spatial ability. *Gifted Child Quarterly*, *42*, 157–171.

Stumpf, H. (2006). *Spatial Test Battery manual*. Baltimore, MD: Johns Hopkins University, Johns Hopkins Center for Talented Youth.

Stumpf, H., & Eliot, J. (1995). Gender-related differences in spatial ability and the *k* factor of general spatial ability in a population of academically talented students. *Personality and Individual Differences*, *19*, 33–45.

Stumpf, H., & Eliot, J. (1999). A structural analysis of visual spatial ability in academically talented students. *Learning and Individual Differences*, *11*, 137–151.

Stumpf, H., & Haldimann, M. (1997). Spatial ability and academic success of sixth grade students at international schools. *School Psychology International*, *18*, 149–163.

Sturzebecher, K. (1972). Raumvorstellung—ein bedeutsamer Intelligenzfaktor in der Schule [Spatial ability—an important factor of intelligence in the schools]. *Die deutsche Schule*, *64*, 690–701.

Super, D. E., & Bachrach, P. B. (1957). *Scientific careers and vocational development theory*. New York, NY: Columbia University, Teachers College Bureau of Publications.

Tendick, F., Downes, M., Goktekin, T., Cavusoglu, M. C., Feygin, D., Wu, X., . . . Way, L. W. (2000). A virtual testbed for training laparoscopic surgical skills. *Presence*, *9*, 236–255.

Thorndike, R. L. (1947). *Research problems and techniques* (AAF Aviation Psychology Research Report No. 3). Washington, DC: Government Printing Office.

Touron, J. (Ed.). (2005). The Center for Talented Youth model [Special issue]. *High Ability Studies*, *16*(1).

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, *117*, 250–270.

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over fifty years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, *101*, 817–835.

Webb, R. M., Lubinski, D., & Benbow, C. P. (2007). Spatial ability: A neglected dimension in talent searches for intellectually precocious youth. *Journal of Educational Psychology*, *99*, 397–420.

Willingham, W. W., & Cole, N. S. (1997). Research on gender differences. In W. W. Willingham & N. S. Cole (Eds.), *Gender and fair assessment* (pp. 17–54). Mahwah, NJ: Lawrence Erlbaum.

Willingham, W. W., Cole, N. S., Lewis, C., & Leung, S. W. (1997). Test performance. In W. W. Willingham & N. S. Cole (Eds.), *Gender and fair assessment* (pp. 55–126). Mahwah, NJ: Lawrence Erlbaum.

AUTHOR BIOS

Heinrich Stumpf, Dipl. Psych., Dr. Phil., studied philosophy and psychology at the University of Bonn, Germany. Since 1975, he has been working in the fields of test construction, test evaluation, psychometrics, and research on individual differences for the University of Bonn, the German National Scholarship Foundation, and Johns Hopkins University. E-mail: heinrich.stumpf1@web.de

Carol J. Mills, PhD, recently retired from her position as Senior Director of Research and Counseling Services for CTY after 24 years. Her research interests include twice-exceptional students, spatial ability, and international applications of above-level specific ability testing. She continues to consult with CTY and to conduct psycho-educational assessments for CTY's Diagnostic and Counseling Center. E-mail: cjmills@jhu.edu

Linda E. Brody, EdD, directs the Study of Exceptional Talent at CTY, where she counsels families on educational options and conducts research on talent development. Her primary research interests focus on evaluating strategies to serve gifted students, such as acceleration, and on special populations, including the highly gifted, gifted females, and twice-exceptional students. E-mail: lbrody@jhu.edu

Philip G. Baxley, JD, is a psychometrician at CTY where he revises and constructs new forms of the School and College Ability Test (SCAT) and assists Dr. Stumpf with the STB revisions. He is the primary technical liaison with Prometric, CTY's testing contractor. He advises CTY, foreign governments, schools, educators, and parents on topics such as test development, above-grade-level testing, and test score interpretation. He earned a Juris Doctor in 1996 from the University of Baltimore School of Law. E-mail: pbaxley@jhu.edu