Psychological Test and Assessment Modeling, Volume 55, 2013 (3), 316-329

Creative Scientific Ability Test (C-SAT): A new measure of scientific creativity

Ugur Sak[1] *& M. Bahadir Ayas*[2]

Abstract

The assessment of creativity has been a controversial issue in the studies of creativity. Contrary to old paradigms, contemporary researchers support the use of domain-specific tests to measure creativity. The purpose of this study was to investigate some psychometric properties of the Creative Scientific Ability Test (C-SAT), a domain-specific test of scientific creativity. The C-SAT was developed based on the Scientific Discovery as Dual Search model and pioneering works on divergent thinking. The test is composed of five subtests and measures fluency, flexibility and creativity and hypothesis generation, hypothesis testing and evidence evaluation in five areas of science. In the study, the C-SAT was administered to 288 sixth grade students in a city in the mid part of Turkey. Factor validity analysis revealed the presence of one component and concurrent validity analysis showed that mathematically talented students scored significantly higher on the C-SAT than did average students. Reliability values of the C-SAT ranged from good (.85) to excellent (.96) and all of the item discrimination correlations were medium or large. Research findings show that the C-SAT can be used as an objective measure of scientific creativity.

Key words: Creative Scientific Ability Test, scientific creativity, assessment

---

[1] *Correspondence concerning this article should be addressed to:* Prof. Ugur Sak, Gifted Education Division, Faculty of Education, Anadolu University, 26470 Eskisehir, Turkey; email: usak@ anadolu.edu.tr

[2] Dept. of Special Education, Faculty of Education, Anadolu University

Introduction

The assessment of creativity involves a number of criteria such as novelty, relevance, effectiveness, usefulness, and surprise (Boden, 2004; Cropley, 1999; Plucker, Beghetto, & Dow, 2004; Sternberg & Lubart, 1995). Inspired by definitions of general creativity, scientific creativity may be defined simply as the ability to generate novel and useful ideas or products. According to this definition, any scientific idea that is extremely original but not useful at all cannot be considered to be creative or vice versa. Thus, a scientific idea needs to possess some degree of originality and usefulness to be accepted as creative. Degrees of originality and usefulness determine the level of creativity of the idea.

Research on the assessment of creativity has been criticized for not having adequate criterion measures as well as for relying on subjective judgments and for using creativity tests that have theoretically too general or unimportant items to measure such a multidimensional complex construct (Baer, 1994; Frederiksen & Ward, 1978; Hocevar, 1979; Kaufman, Plucker, & Baer, 2008). Baer (1994), Lubart and Guignard (2004) and Kaufman, Plucker, and Baer (2008) further criticized the inefficiency of general creativity tests in predicting real creativity and suggested the use of domain-specific tests of creative ability. Baer (1991, 1993), for example, carried out a series of research studies with fourth and eighth grade students and adults to investigate generality and specificity of creativity, using poems, stories, equations, mathematical word problems and collages tasks. The results of these studies showed a range of correlations from -0.05 to 0.08, supporting a task-specific view of creativity. Based on these findings, Baer (1994) argued that creative performance on one task is not predictive of creative performance on other tasks, including those in the same domain. Other researchers also criticized the use of the generality approach in the measurement of creativity and supported the use of domain-specific tests (Frederiksen & Ward, 1978; Hu & Adey, 2002; Kaufman & Baer, 2002; Runco, 1987; Plucker, 1998).

Motivated by the studies and the ideas discussed above and a strong need for a domain-specific test of scientific creativity, we developed the Creative Scientific Ability Test to measure scientific creativity of students and to identify scientifically creative students (Ayas & Sak, 2008, 2009; Ayas, 2010; Sak, 2010). We first review scientific creativity, and then, the Creative Scientific Ability Test for measuring creative potential in students in the sciences.

Scientific creativity

Research on scientific creativity shows that scientific creativity can be viewed as a process of an interaction among general-creativity skills, science-related skills and scientific knowledge in one or more areas of science and motivation, interest, concentration and search for knowledge and chance permutation of mental elements (Dunbar, 1999; Heller, 2007; Klahr, 2000; Puccio, 1991; Roe, 1952, 1961; Simonton, 1988; Subotnik, 1993; Torrance, 1992). Scientific giftedness is even viewed as a convergence of artistic and scientific abilities (Innamorato, 1998). The process of knowledge construction in science makes it different from other disciplines in that creative scientific ideas are built on a vast theoretical, technical and experimental knowledge (Dunbar, 1999). More specifically, scientific creativity involves a whole array of activities such as designing and performing experiments, inferring theories from data, modifying theories, inventing instruments (Kulkarni & Simon, 1988), formulating hypotheses, solving problems (Klahr, 2000; Klahr & Dunbar,1988; Newell & Simon, 1972; Simon, 1977), working on the unexpected (Dunbar, 1993) and on opposite ideas (Rothenberg, 1971, 1996).

Simon (1977) proposed that scientific creativity can be thought of as problem solving involving search in various problem spaces. In a similar line of thinking, Klahr (2000) proposed a two-dimensional taxonomy of the major components of scientific creativity. One dimension represents domain-specific and domain-general knowledge. The other dimension includes major processes involved in scientific discoveries: generating hypotheses, designing experiments, and evaluating evidence. The acquisition of domain-specific knowledge is important because it influences not only knowledge structure in the domain but also the processes used in the generation and evaluation of new hypotheses in that domain. During scientific discovery, domain-specific knowledge and domain-general heuristics guide the design of experiments, selection or formulation of new hypotheses and verification of evidence.

Scientific creativity as dual search

Klahr and Dunbar (1988) put forward that scientific creativity includes two primary spaces: a space of hypotheses and a space of experiments. In search of the two spaces, they identified two different strategies for generating new hypotheses: (a) searching memory for possible hypotheses (hypothesis space) and (b) conducting experiments until a new hypothesis could be generated from data (experiment space). Using this framework, Klahr (2000) proposed that scientific discovery could be understood as a dual search (SDDS): search in hypothesis space and search in experiment space. The SDDS consists of basic components that guide within and between the two spaces. Problems to be solved in each space, as well as spaces themselves are rather different from each other in that they require different representations and operators. Search in two spaces requires three major processes that work interdependently (Klahr): hypothesis space search, experiment space search, and evidence evaluation. These three processes guide the entire process of scientific creativity from formulation of hypotheses, through experimental evaluations to decisions to accept or reject hypotheses.

Hypothesis space. The initial stage of a hypothesis space consists of some knowledge while the goal stage is a hypothesis that could explain that knowledge (Klahr, 2000). After hypotheses are generated, they are verified for their plausibility. New hypotheses originate from two different sources. One of the sources consists of prior knowledge. The other source arises from experimental data. The output of search in hypothesis space is a completely specified hypothesis. This new hypothesis becomes an input for experiment design.

Experiment space. Scientists search in the experiment space to design and carry out appropriate experiments that could prove or disprove their hypotheses. According to the SDDS, the process of testing a hypothesis involves designing an experiment appropriate to the hypothesis, making predictions, running the experiment, and matching the outcomes of the experiments to the predictions (Klahr, 2000). Testing a hypothesis produces evidence in experiment space for or against the hypothesis under testing. This evidence is used as an input in the evidence evaluation process.

Evidence evaluation. In the evidence evaluation process, predictions articulated in hypotheses are compared with results obtained in experiments (Klahr, 2000). Evidence evaluation determines whether evidence obtained from experiments is sufficient to accept or reject hypotheses. If evidence is not sufficient, the process restarts from the hypothesis space or hypothesis testing. That is, the evidence evaluation process mediates search in both hypothesis space and experiment space and assesses the fit between the theory and evidence.

According to the SDDS, scientific creativity involves an interaction of hypothesis generation, experiment design, and evidence evaluation processes. These three processes coordinate the search in hypothesis space and experiment space. Furthermore, Dunbar (1993) found that one of the key aspects of making a scientific discovery is to switch goals from testing a favored hypothesis to a goal of accounting for unexpected findings. Being able to focus on unexpected findings is a key component of scientific discovery. On the other hand, focusing only on one hypothesis, while ignoring other possible hypotheses, can produce a type of creativity block. This type of thinking can mislead the design of experiments, formulation of theories and interpretation of data (Dunbar, 1999). That is, the ability to produce many hypotheses for a problem situation can be one of the key characteristics of scientific creativity.

The Creative Scientific Ability Test (C-SAT)

The Scientific Discovery as Dual Search model (Klahr, 2000; Klahr & Dunbar, 1988) and pioneering works on divergent thinking and its measurement (Guilford, 1950, 1956, 1967; Torrance, 1962, 1988) provided a general theoretical framework for the development of the C-SAT (Ayas & Sak, 2008; 2009; Ayas, 2010; Sak, 2010). It has three dimensions, with the creativity process on a dimension, the scientific process on another dimension, and the areas of science on the third dimension (see table 1). Based on this framework, fluency, flexibility and composite creativity were identified as the main components of general creativity. Contrary to widespread practices in the assessment of creativity, we did not include a distinct ideational originality in the C-SAT because originality has been found to have very high correlations with fluency (Mouchiroud & Lubart, 2001). Hypothesis generation, experiment design and evidence evaluation were determined as the major components of scientific creativity. Further, five areas of science (biology, physics, chemistry, ecology, and interdisciplinary science) were identified. We limited the number of skills used in the framework to the theoretically most

Table 1:
The Theoretical Structure of the Creative Scientific Ability Test

Subtests	Domain		
	Areas of Science	**Scientific Process**	**Creativity Process**
Fly Experiment	Biology	Hypothesis Generation	Fluency, Flexibility, Creativity
Interaction Graph	Interdisciplinary science	Hypothesis Generation	Fluency, Flexibility, Creativity
Sugar Experiment	Chemistry	Hypothesis Testing	Fluency, Flexibility, Creativity
String Experiment	Physics	Hypothesis Testing	Fluency, Flexibility, Creativity
Food Chain	Ecology	Evidence Evaluation	Fluency, Flexibility, Creativity

important ones – otherwise, the assessment of scientific creativity would be practically impossible – because, both general creativity and scientific creativity involve many thinking and problem solving processes. That is, any single measurement of scientific creativity can necessarily focus only on some aspects of the whole process.

The Subtests of the C-SAT

The C-SAT measures potential for scientific creativity in sixth through eighth grade students. It specifically measures fluency, flexibility and creativity in hypothesis generation, experiment design and evidence evaluation tasks in five areas of science. It includes five subtests with one subtest from each area (table 1): fly experiment (biology), change graph (interdisciplinary science), sugar experiment (chemistry), string experiment (physics), and food chain (ecology). Each subtest consists of one open-ended problem. Two subtests include hypothesis generation problems, two subtests contain experiment design problems, and one subtest includes an evidence evaluation problem.

Subtest 1: fly experiment. This problem presents a figure of an experiment designed by a researcher. Students are required to generate as many hypotheses as they can think of that the researcher might want to test by this experiment. The purpose of this task is to measure fluency, flexibility and creativity in hypothesis generation in the area of biology.

Subtest 2: change graph. This problem presents a graph of reverse changes in the amounts of two variables and an effect that starts these changes. Students are asked to think of as many pairs of variables as they can that fit the graph. This task measures fluency, flexibility, and creativity in hypothesis generation in interdisciplinary science.

Subtest 3: sugar experiment. A figure of an experiment designed by a researcher and a graph showing the researcher's hypothesis are presented in this problem. Students are required to think of as many changes as they can that should be made in the experiment in order for the researcher to prove the hypothesis. This task measures fluency, flexibility, and creativity in experiment design (hypothesis testing) in the area of chemistry.

Subtest 4: string experiment. A figure of an experiment is presented in this problem. Students are asked to think of as many changes as they can that should be made in the experiment to achieve a goal. The purpose of this task is to measure fluency, flexibility, and creativity in experiment design (hypothesis testing) in the area of physics.

Subtest 5: food chain. This problem presents a figure of a food chain and a graph of the change in this food chain. Students are asked to think of as many causes as they can of the change. This problem measures fluency, flexibility, and creativity in evidence evaluation in the area of ecology.

The-C-SAT yields a fluency, flexibility, and creativity score for each subtest and a total fluency, total flexibility and total creativity score for the total test. Fluency is the number of correct responses generated for each task. Flexibility is the number of conceptual categories among responses. Contrary to traditional practices that sum fluency and flexibility scores to calculate the total creativity, the C-SAT creativity is estimated using the Creativity Quotient (CQ) formula (log2) proposed by Snyder, Mitchell, Bossomaier and Pallier (2004, p. 416): $CQ = log_2 \{(1+ u_1) (1+ u_2) \dots (1+ u_c)\}$. In the formula, “u” denotes the number of correct responses in a distinct category. This innovative method promotes flexibility by increasing its contribution to the total creativity score. A strong belief is that ideas in distinctly different categories should be weighted more than those that fall in the same category (Getzels & Jackson, 1962; Guilford, 1959). The fluency and flexibility scores are used for research purposes, whereas the composite creativity score can be used both for research purposes and for the identification of scientifically creative students.

We carried out a series of research studies on the psychometric properties of the C-SAT (Ayas & Sak, 2008, 2009; Ayas, 2010; Sak, 2010). We found in these studies that the C-SAT had strong criterion validity, with medium to high medium correlations between students’ scores on the C-SAT and their science and math grades and their scores on other math and science related measures. A factor analysis on the C-SAT scores yielded one-factor solution. Its various types of reliability ranged from 0.86 to 0.97. That is, previous studies yielded strong validity and reliability evidence for the C-SAT. However, after five years experience of using and working on the C-SAT, several revisions were made on the instructions and on some specifics of the items to reasonably limit the possible correct responses that could be generated for each item. In the current study, we attempted to replicate previous studies and revalidate the C-SAT after its last revision. We were particularly interested if the revised C-SAT would show similar psychometric properties to those we obtained in previous studies.

Method

Participants

The study included 288 sixth-grade students who applied to the Education Programs for Talented Students (EPTS), an after-school program for talented students, at Anadolu University in the City of Eskisehir in Turkey. They came from 52 different schools from 4 cities in the mid part of Turkey. Female participants comprised 52.7% of the total sample; male participants comprised 47.3%. Application to the EPTS was free and no criteria were used as a precondition for the application; therefore, any student who was willing to participate in the program had the choice to apply. Because the EPTS offers advanced courses for talented students, most of the students who apply to the EPTS have high ability and high academic achievement in mathematics and the sciences.

Instrument

The Test of Mathematical Talent (TMT). In the study, besides the Creative Scientific Ability Test, the Test of Mathematical Talent was used to collect data. The purpose of the TMT is to identify sixth grade students who have high ability in mathematics. It is a multiple-choice, paper-and-pencil test. Its psychometric properties were investigated in a number of studies and it was found to have good validity and reliability evidence (Sak, Turkan, Sengil, Akar, Demirel, & Gucyeter, 2009; Sengil, 2009). The test has a mean of 100 and a standard deviation of 15. Students' scores on the TMT in the current study ranged from 64.59 to 140.74, with a mean of 100. The group was found to be heterogeneous in terms of their mathematical ability.

Data collection

The Test of Mathematical Talent (TMT) and the C-SAT are the two identification instruments of the EPTS. Students who apply to the EPTS are required to take the two tests. In this study, the tests were administered to all the students in two sessions on the University's Campus in the same day. First, the TMT was administered to the students. This session lasted 80 minutes. After a 40-minute break, the students took the C-SAT. This session lasted 40 minutes. The TMT answers were scored by a computer program. The C-SAT answers were scored by two trained scorers using the C-SAT standard scoring procedures.

Results

The results included item-level descriptions and reliability and validity of the C-SAT. First, an item level analysis was carried out to determine minimum and maximum scores and means and standard deviations for the fluency, flexibility and creativity in each subtest as well as for the total fluency, total flexibility and total creativity. Students' scores ranged from 0 to 29 (M= 10.68; SD= 6,16) for the total fluency, 0 to 15 (M= 6,56; SD= 3,22) for the total flexibility and 0 to 21 (M= 8,60; SD= 4,52) for the total creativity. That is, the highest scoring students had 29 correct responses for all the subtests and generated 15 distinct categories among correct responses.

Item discrimination

First, correlation coefficients between the items and the total creativity score and corrected item-total correlations were calculated (see table 2). Corrected item-total correlations ranged from medium to large. The subtest 4 fluency score had the smallest corrected correlation with the total (.385), whereas the subtest 2 flexibility had the largest corrected correlation with the total (.578). Second, the students were classified as high (top 27%), middle (middle 46%) and low ability (bottom 27%) based on their total creativity scores on the C-SAT to compare their performance on the subtests. Then, a one-way between-groups multivariate analysis of variance (MANOVA) was carried out to exam-

Table 2:
Item-Total Correlations

Subtest	Subscore	Item-Total Correlation*	Corrected Item-Total Correlation*
Subtest 1	Fluency	.539	.433
	Flexibility	.47	.408
	Creativity	.538	.466
Subtest 2	Fluency	.569	.467
	Flexibility	.523	.459
	Creativity	.577	.504
Subtest 3	Fluency	.634	.519
	Flexibility	.648	.578
	Creativity	.651	.561
Subtest 4	Fluency	.494	.385
	Flexibility	.497	.425
	Creativity	.505	.415
Subtest 5	Fluency	.545	.518
	Flexibility	.531	.461
	Creativity	.631	.548

Note. All of the correlations are significant at the 0.01 level (2-tailed).

ine subtest discriminations. Dependent variables (five subtests) typically were combined in the multivariate analyses. The preliminary analysis indicated an inequality of error variance between the variables. Therefore, the alpha level for statistical significance was set at .01, as recommended by Tabachnick and Fidell (1996). The multivariate results showed that the three groups differed significantly on the combined dependent variables: [$F(10, 558) = 78.5$, $p < .001$; Wilks' Lambda= .17; partial eta squared = .59]. After finding a significant difference in the MANOVA, post hoc tests were used to explore significant differences among the groups. Follow-up pairwise comparisons showed that the high ability group scored significantly higher than the middle and the bottom group on all the subtests and the middle group scored significantly higher than the bottom group on all the subtests. All of the significance values were smaller than .01.

Reliability

Because student responses on the C-SAT are scored by trained scorers using the standard scoring procedures of the C-SAT, an analysis was carried out to examine its inter-scorer reliability. Eighty-three cases (28%) of the 288 participants were randomly selected and their papers were scored by two independent scorers who were trained in the C-SAT scoring procedures. The analysis showed that the interscorer reliability coefficient was .96 for the total fluency subtest, .94 for the total flexibility subtest, and .96 for the total creativity subtest. In the second reliability analysis, the C-SAT's internal consistency reliability was examined. The Cronbach's Alpha coefficient value was found to be .848. In sum, reliability coefficients of the C-SAT ranged from good to excellent.

Construct validity

First, the suitability of data for factor analysis was assessed. The Kaiser-Meyer-Oklin value was found to be .66, exceeding the recommended value of .6. The Barlett's Test of Sphericity reached statistical significance, supporting the factorability of the correlation matrix. Because the fluency, flexibility and creativity subscores of each subtest of the C-SAT are obtained from the same set of responses, instead of the subscores, the five subtests of the C-SAT were subjected to Principal Components Analysis (PCA). The analysis revealed one component with eigenvalue exceeding 1. One factor-solution explained 34.45% of the variance. Factor loadings ranged from .47 to .63. An inspection of the screeplot revealed a clear break after the first component. In addition, the Parallel Analysis showed evidence of one factor.

Concurrent validity

The concurrent validity of the C-SAT was examined by comparing students' performance on the C-SAT and the Test of Mathematical Ability (TMT) using the independent samples t-test analysis. First, students were classified into two groups based on their scores on the TMT. Those who scored in the top 10% on the TMT were classified as the high mathematical ability group and the remaining 90% were classified as the average mathematical ability group. Because the C-SAT total fluency, total flexibility and total creativity scores correlated with each other very highly, only the total creativity score was used as a dependent variable. The Levene's test for equality of variances was used to check the equality of variance for the groups. The analysis showed that the equal variance assumption was not violated (p= .776). The results of the independent samples t-test showed a significant difference between the means of the two groups [M1= 12.60, SD= 4.62; M2=8.16, SD=4.29; $t(286)=5.23$, $p<.000$]. That is, the students who had high mathematical ability scored significantly higher on the C-SAT than did the students who had average mathematical ability. The effect size of the difference between the means was found to be .995, a large effect (Cohen, 1988).

Discussion and conclusions

In the current study, an investigation was carried out to explore some psychometric properties of the revised C-SAT. The study included an examination of the reliability, validity and item descriptions of the C-SAT. The reliability and validity findings ranged from good to excellent.

One of the most serious drawbacks of most creativity tests currently in use is related to the nature of items and their difficulty levels. One can generate an unlimited number of responses for a single item in many of these tests, such as uses of objects and making changes in objects. Because of this drawback, scoring responses for these tests has been found to be labor intensive and impractical (Kaufman, Plucker, & Baer, 2008; Kim, 2006). Indeed, this problem is one of the major barriers for the widespread use of creativity tests in the identification of gifted and creative students. The findings obtained in the current study show that the C-SAT items are moderately open-ended and reasonably difficult. That is, the most creative students had a limited number of correct responses for all the subtests. However, the limited open-endedness of the problems should not be interpreted as a limitation of the C-SAT. Because, although the problems are open-ended to some degree, even the most creative students could not generate all the possible correct responses for each subtest. In the sample of the current study, the most creative students had 29 correct responses out of 267 possible correct responses. These findings provide strong evidence for the practicality of the C-SAT scoring and also show that it does not have a ceiling effect.

We hypothesized that mathematically talented students would show higher performance on the C-SAT than would students with average mathematical ability, because scientific ability and mathematical ability and science and mathematics are considered to be theoretically related (Cain & Lee, 2010; Davis, 2009; Delialioglu & Askar, 1999; Heller, 2007; Li, Shavelson, Kupermintz, & Ruiz-Primo, 2002). In fact, scientists usually use mathematics as a tool in their studies. There is even a science discipline called mathematical physics. The current study supported the hypothesis and yielded strong evidence for the concurrent validity of the C-SAT. According to the finding, students who were better on mathematical ability also were better on scientific creativity and their performance was reflected on the C-SAT. Moreover, as a support to the validity of the C-SAT, its predictive validity to discriminate between ability groups in scientific creativity should be investigated in future studies.

In conclusion, research findings obtained in previous studies as well as in this study support the use of the C-SAT as an objective measure of scientific creativity in sixth grade students. It can be used for two purposes: (1) as a criterion in research and (2) as a supplement in the identification of creatively gifted students in the sciences. Related to the former purpose, it can be used to examine effects of creativity training programs in scientific creativity, to investigate gender, age and grade differences in scientific creativity, to study scientific creativity in cross cultural settings, to examine the development of scientific creativity in longitudinal studies, to explore creativity in general and domain-specific forms, to examine the relationship between process and product in scientific creativity, to study students' performance in hypothesis generation, hypothesis testing and evidence evaluation tasks, and to investigate relationships of scientific creativity with other variables.

The C-SAT also can be used in the identification of gifted students. IQ and achievement tests are pioneering instruments in the identification of gifted students, and therefore have predominated in identification practices (Davis, Rimm, & Siegle, 2011; Feldhusen, Asher, & Hoover, 1984; Johnsen, 2004; Renzulli & Delcourt, 1986). Although the contribution of these tests to the field of gifted education is exceptionally valuable, the widespread and even unquestionable use of these tests by practitioners has challenged the development and widespread use of other types of identification instruments. IQ and achievement type of identification can identify high-ability students in different areas but can miss many students who have high creative potential in specific domains. Torrance (1962), for example, found that creatively gifted students missing such points as 130 IQs achieved as well as their classmates with IQs in excess of 130 who would be classified as creatively gifted by similar standards. He further stated that the identification of gifted students only on the basis of IQ and scholastic aptitude tests could eliminate about 70% of the top 20% of creative students from consideration. This limitation of IQ and achievement oriented practices have led researchers to recommend and to use multiple criteria (Frasier, 1997). The C-SAT can be used in combination with other instruments in sample-based identification practices as one of multiple criteria to identify creatively gifted students in science. It also can be used as an additional source of information in norm-based identifications.

Acknowledgement

This research was partly supported by a research grant (#1003E76) from the Anadolu University Scientific Projects Office.

References

Ayas, B. (2010). *An investigation of the psychometric properties of the Creative Scientific Ability Test.* Unpublished master's thesis. Anadolu University, Eskisehir, Turkey.

Ayas, B., & Sak, U. (2008). The test of scientific creativity: Its development and psychometric properties. *Paper presented at the 4th International Conference on Intelligence and Creativity,* Munster, Germany.

Ayas, B., & Sak, U. (2009). Bilimsel Üretkenlik Testi: Teorik alt yapısı, geliştirilme süreci ve psikometrik özelikleri. *Paper presented at the 2nd National Conference on Talented Children,* Eskisehir, Turkey.

Baer, J. (1991). Generality of creativity across performance domains. *Creativity Research Journal, 4,* 23-39.

Baer, J. (1993). *Creativity and divergent thinking: A task-specific approach.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Baer, J. (1994). Why you shouldn't trust creativity tests. *Educational Leadership, 51,* 80-83.

Boden, M. A. (2004). *The creative mind: Myths and mechanisms.* London: Routledge.

Cain, R. W., & Lee, E. C. (2010). An analysis of the relationship between science and mathematics at the secondary school level. *School Science and Mathematics, 63,* 9, 705-713.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Cropley, A. (1999). Definitions of creativity. In S. R. Pritzker & M. A. Runco (Eds.), *Encyclopedia of creativity,* (vol. 2., pp. 511-524). San Diego, CA: Academic Press.

Davis, G. A., Rimm, S. B., & Siegle, D. (2011). *Education of the gifted and talented.* 6th ed. New Jersey: Pearson.

Davis, C. A., (2009). Gender, mathematics, reading comprehension and science reasoning as predictors of science achievement among African-American students at a historically black college or university. *Unpublished doctoral dissertation,* University of Southern Mississippi, Mississippi.

Delialioglu, O., & Asker, P. (1999). Contribution of students' mathematical skills and spatial ability to achievement in secondary school physics. *Hacettepe University Journal of Education, 16,*17, 34-39.

Dunbar, K. (1993). Concept discovery in a scientific domain. *Cognitive Science, 17,* 397-434.

Dunbar, K. (1999). Science. In M. A. Runco & S. R. Pritzker (Eds), *Encyclopedia of creativity,* vol 2, (pp. 525-531). San Diego, CA: Academic Press.

Feldhusen, J. F., Asher, J. W., & Hoover, S. M. (1984). Problems in the identification of giftedness, talent, or ability. *Gifted Child Quarterly, 4,* 149-151.

Frasier, M. M. (1997). Multiple criteria: The mandate and the challenge. *Roeper Review, 20,* 2, A4-A6.

Frederiksen, N., & Ward, W. C. (1978). Measures for the study of creativity in scientific problem solving. *Applied Psychological Measurement, 2,* 1, 1-24.

Getzels, J. W., & Jackson, P. W. (1962). *Creativity and intelligence.* New York: John Wiley and Sons, Inc.

Guilford, J. P. (1950) Creativity. *American Psychologist, 5,* 444–454.

Guilford, J. P. (1956) The structure of intellect. *Psychological Bulletin, 53,* 267–293.

Guilford, J. P. (1959). *Personality.* New York: McGraw Hill.

Guilford, J. P. (1967). *The nature of human intelligence.* NewYork: McGraw-Hill.

Heller, K. A. (2007). Scientific ability and creativity. *High Ability Studies, 2,* 209-234.

Hocevar, D. (1979). A comparison of statistical infrequency and subjective judgment as criteria in the measurement of originality. *Journal of Personality Assessment, 3,* 297-299.

Hu, W. & Adey, P., (2002). A test of scientific creativity for secondary school students. *International Journal of Science Education, 4,* 389-403.

Innamorato, G. (1998). Creativity in the development of scientific giftedness: Educational implications. *Roeper Review, 1,* 54-60.

Johnsen, S. K. (2004). *Identifying gifted students: A practical guide.* Waco, TX: Prufrock Press.

Kaufman, J. C., & Baer, J. (2002). Could Steven Spielberg manage the Yankees?: Creative thinking in different domains. *The Korean Journal of Thinking and Problem Solving, 12*(2), 5-14.

Kaufman, J. C., Plucker, J. A., & Baer, J. (2008). *Essentials of creativity assessment.* New Jersey: John Wiley & Sons, Inc.

Kim, K. H. (2006). Can we trust creativity tests? A review of the Torrance Tests of Creative thinking (TTCT). *Creativity Research Journal, 1,* 3-14.

Klahr, D. (2000). *Exploring science: The cognition and development of discovery processes.* Cambridge: The MIT Press.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12,* 1-48.

Kulkarni, D., & Simon, H. A. (1988). The processes of scientific theories: The strategy of experimentation. *Cognitive Science, 12,* 139-175.

Li, M., Shavelson, R. J., Kupermintz, H., & Ruiz-Primo, M. A. (2002). On the relationship between mathematics and science achievement in the United States. In D. F. Robitaille & A. E. Baeton (Eds.), *Secondary analysis of the TIMMS data* (pp. 233-249). Netherlands: Kluwer Academic Publishers.

Lubart, T., & Guignard, J. H. (2004). The generality-specificity of creativity: A multivariate approach. In R. J. Sternberg, E. Grigorenko & J. L. Singer (Eds.), *Creativity: From potential to realization* (pp. 43−56). Washington: A.P.A..

Mouchiroud, C., & Lubart, T. (2001). Children's original thinking: An empirical examination of alternative measures derived from divergent thinking tasks. *The Journal of Genetic Psychology, 4,* 882-401.

Newell, A., & Simon, H. A. (1972). *Human Problem Solving.* Englewood Cliffs, N.J.: Prentice-Hall.

Plucker, J. A. (1998). Beware of simple conclusions: The case for the content generality of creativity. *Creativity Research Journal, 11,* 179-182.

Plucker, J., Beghetto, R. A., & Dow, G. (2004). Why isn't creativity more important to educational psychologists? Potential, pitfalls, and future directions in creativity research. *Educational Psychologist, 39,* 83–96.

Puccio, G. J. (1991). William Duff's Eighteenth Century examination of original genius and its relationship to contemporary creativity research. *Journal of Creative Behaviour, 25*(1), 1-10.

Renzulli, J. S., & Delcourt, M. A. B. (1986). The legacy and logic of research on the identification of gifted persons. *Gifted Child Quarterly, 1,* 20-23.

Roe, A. (1961). The psychology of a scientist. *Science, t34,* 56-459.

Roe, A. (1952). A psychologist examines 64 eminent scientists. *Scientific American, 185*(5), 21-25.

Rothenberg, A. (1971). The process of janusian thinking in creativity. *Archives of General Psychiatry, 24,* 195-205.

Rothenberg, A. (1996). The janusian process in scientific creativity. *Creativity Research Journal, 9,* 207-209.

Runco, M. A. (1987). The generality of creative performance in gifted and nongifted children. *Gifted Child Quarterly, 31*, 121-125.

Sak, U. (2010). Assessment of creativity: Focus on math and science. *Paper presented at the 12th ECHA Conference*, Paris, France.

Sak, U., Turkan, Y., Sengil, S., Akar, A., Demirel, S., & Gucyeter, S. (2009). Matematiksel Yetenek Testi (MYT)'nin gelişimi ve psikometrik özellikleri (Development and psychometric properties of the Test of Mathematical Talent). *Paper presented at the 2nd National Conference on Talented Children,* Eskisehir, Turkey.

Sengil, S. (2009). İlköğretim 6. ve 7. sınıf öğrencilerine yönelik matematik yetenek testi'nin kapsam geçerliği (Content validity of the Test of Mathematical Talent). *Unpublished master's thesis,* Anadolu University, Eskisehir, Turkey.

Simon, H.A. (1977). *Models of discovery.* Dordrecht: Reidel.

Simonton, D.K. (1988). *Scientific genius: A psychology of science.* Cambridge, England: Cambridge University Press.

Snyder, A., Mitchell, J., Bossomaier, T., & Pallier, G. (2004) The Creatvity Quotient: An objective scoring of ideational fluency. *Creativity Research Journal, 16*(4): 415-420.

Subotnik, R. F. (1993). Adult manifestations of adolescent talent in science. *Roeper Review, 15*(3), 164-169.

Sternberg, R. J., Lubart, T. I. (1995). *Defying the crow: Cultivating creativity in a culture of conformity.* New York: The Free Press.

Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics.* 2nd ed. New York: HarperCollins.

Torrance, E. P. (1962). *Guiding creative talent.* Englewood Cliffs, NJ: Prentice-Hall.

Torrance, E. P. (1988). The nature of creativity as manifest in its testing. In R. J. Sternberg (Ed.), *The nature of creativity* (pp. 43-75). New York: Cambridge Univ. Press.

Torrance, E. P. (1992). A national climate for creativity and invention. *Gifted Child Today, 15*(1), 10-14.