JSGE
The Journal of Secondary Gifted Education
Vol. XVI, No. 2/3, Winter/Spring 2005, pp. 57–66.

Can Only Intelligent People Be Creative?

A Meta-Analysis

Kyung Hee Kim
Eastern Michigan University

Some research has shown that creativity test scores are independent from IQ scores, whereas other research has shown a relationship between the two. To clarify the cumulative evidence in this field, a quantitative review of the relationship between creativity test scores and IQ scores was conducted. Moderating influences of IQ tests, IQ score levels, creativity tests, creativity subscales, creativity test types, gender, age, and below and above the threshold (IQ 120) were examined. Four hundred forty-seven correlation coefficients from 21 studies and 45,880 participants were retrieved. The mean correlation coefficient was small (r = .174; 95% CI = .165 – .183), but heterogeneous; this correlation coefficient indicates that the relationship between creativity test scores and IQ scores is negligible. Age contributed to the relationship between intelligence and creativity the most; different creativity tests contributed to it secondly. This study does not support threshold theory.

Guilford (1962) hypothesized that creative individuals possess divergent thinking abilities that traditional IQ tests do not measure, and other researchers have shown that creativity test scores, divergent thinking tasks, and creative achievement are independent from IQ (e.g., Getzels & Jackson, 1958; Gough, 1976; Guilford, 1950; Helson & Crutchfield, 1970; Helson & Crutchfield 1971; Herr, Moore, & Hansen, 1965; Rossman & Horn, 1972; Rotter, Langland & Berger, 1971; Torrance, 1977). In seeming contrast, other research has shown a relationship between creativity test scores and IQ scores (e.g., Runco & Albert, 1986; Wallach, 1970).

However, many researchers agree with the threshold theory, which explains that creativity and intelligence are separate constructs; that is, more intelligence does not necessarily mean greater creativity. Threshold theory assumes that, below a critical IQ level (which is usually said to be about 120), there is some correlation between IQ and creative potential, and above it there is not (Barron, 1961; Getzels & Jackson, 1962; Guilford, 1967; Guilford & Christensen, 1973; MacKinnon, 1961, 1962, 1967; Simonton, 1994; Walberg, 1988; Walberg & Herbig, 1991; Yamamoto, 1964). Although threshold theory supposes that intelligence and creativity are related only up to an IQ of approximately 120, investigations of the threshold theory have contradictory and inconclusive results (e.g., Runco & Albert, 1986). This inconsistency may come from different measures of intelligence and creativity and various features of samples, including gender, age, and socioeconomic status (SES).

Meta-analysis, or quantitative synthesis, attempts to find resolution of apparent conflicts in literature, to discover consistencies and account variability in similarly appearing studies, and to identify core issues for future research (Cooper & Hedges, 1994). Thus, this study synthesized empirical research in the areas of creativity and intelligence for the purpose of creating a generalization about the relationship between creativity and intelligence.

The four primary purposes of this synthesis were to

1. conduct a quantitative synthesis of correlations between IQ and creativity test scores;
2. compare the correlations between IQ and creativity scores for IQ above 120 with those for IQ below 120 to confirm threshold theory;
3. identify some of the variables that moderate those correlations (e.g., IQ tests, different levels of IQ scores, creativity tests, creativity test types, creativity subscales, gender, and age); and
4. use the correlations derived from the quantitative synthesis to investigate models of the relationships between intelligence and creativity.

Methods

Literature Review

More than 100 studies published from 1961 through the summer of 2004 were located from computer searches and personal retrieval within the body of English-language creativity and intelligence literature using Academic Search Premier, ERIC, PsycARTICLES, PsycINFO, and bibliographic searches of each reference. *IQ*, *creativity*, *intelligence*, and *threshold theory* were the keywords used. The criteria for study selection included the reporting of the correlations among measures of intelligence and creativity; *t* tests, chi-square, and *F* tests with a single *df*, as well as exact *p* values. Much of the research examined in the present study failed to report detailed information on procedures, results, or both, thus making it difficult or impossible to generate correlation coefficients. Some studies did report on IQ scores and creativity scores, but the correlation coefficients that were reported were IQ and creativity scores with various creative achievements.

Effect Size Calculations

A quantitative synthesis of the remaining 21 studies was conducted, which was assisted by Schwarzer's Meta 5.3 statistical software. Fisher's z transformation of r was used for analyses in order to adjust for the nonnormal distribution of r. Most researchers believe that studies that employ large samples should get more credit than those that are based on small samples because correlations are known to become more stable as sample size increases (Schwarzer, 1991). Thus, the effect size z_r was weighted by sample size: The weighted mean $z_r = \sum(Nj - 3)\, z_{rj} / \sum(Nj - 3)$, in this equation z_{rj} is the Fisher z_r, corresponding to any r (Rosenthal, 1991). Each mean r was tested using a random effects model of variance, and the reported values of r are back transformations from z (Hunter, Schmidt, & Jackson, 1982). A correlation coefficient was judged heterogeneous when the residual standard deviation exceeded three fourths of the effect size and sampling error was less than 75% of the observed variance (Schwarzer). The residual standard deviation was used as the standard error in estimating 95% confidence intervals (CI).

Four-hundred forty-seven correlation coefficients were retrieved from 21 studies, for a total sample size of 45,880 people. Multiple correlation coefficients were obtained from studies that included more than one correlation using several creativity subscales and reported separate results for gender. When using several correlation coefficients per study in analyses of moderators, a conservative statistical criterion ($p < .001$) was used to protect against Type I error (Rosenthal, 1991; Rosenthal & Rubin, 1984). As reported above, individual correlation coefficients were weighted by sample sizes (Hedges & Olkin, 1985; Hunter & Schmidt, 1990; Rosenthal) in order to adjust for sampling errors in r, giving more credence to studies with large samples because small sample sizes can increase Type II error (Hedges & Olkin).

Moderator Variables

In order to test whether correlation coefficient sizes vary systematically across differing levels of variables that are posited to influence the relationship between intelligence and creativity, analyses of moderator variables were conducted. Correlation coefficient size heterogeneity can be explained through moderator variables. The moderator variables were categorized in a manner that seemed theoretically valid. The correlation coefficients between divergent thinking abilities and intelligence vary widely depending upon the divergent thinking tests, the heterogeneity of the sample, and the testing conditions (Barron & Harrington, 1981). Therefore, factors that might moderate the estimated population correlation coefficients between IQ scores and creativity scores in the present study were considered by comparing correlation coefficients based on various IQ tests, levels of IQ scores, several creativity tests, creativity test types, creativity subscales, gender, age, and IQ scores below and above 120. Most studies did not report information about subjects' ethnicity, making it impossible to include ethnicity as a moderator variable.

When data was entered, the creativity test subscales of associational fluency, ideational fluency, and verbal and figural fluency were classified as "fluency," whereas spontaneous flexibility, adaptive flexibility, and verbal and fig-

-.9 I
-.8 I
-.7 I
-.6 I
-.5 I
-.4 I 168
-.3 I 245688
-.2 I 0001255689
-.1 I 0001111222222222233444466779
-.0 I 0000000000011112223333333333334455666677788889
+.0 I 00000000001111112222222222223344444444455555556666666666777777777777788888888899999999999999
+.1 I 00000000000001111111111122222222222222333333344444444444445555555555666666666666677777788888888888888888899999999999
+.2 I 0000000000000111111111112222222333334444444444555555555666677777777788888889999999
+.3 I 00000001111222233344445556666667777788888999
+.4 I 0000111112222334455667789
+.5 I 00356
+.6 I 0
+.7 I 6
+.8 I
+.9 I

Figure 1. Stem-and-leaf-display for 447 correlation coefficients (*r*) between IQ and creativity scores

Note. The correlation coefficients r ranged from -.46 in the first row of the display to +.76 in the last row.

ural flexibility were classified as "flexibility." Both verbal and figural originality were categorized as "originality." In several of the creativity tests in the present study (e.g., Flescher, 1963; Getzels & Jackson, 1962; Hocevar, 1980; Kogan & Pankove, 1972; Vernon, 1972; Wallach & Kogan, 1965), subscales were not reported. Thus, the tasks in these studies were examined and categorized based on the nature of those tasks. For age, 87 kindergarteners (from Ward, 1968) were included in the elementary group, and naval recruits (from Richards, 1976) and college students were included in the adult group.

Hedges' *g* was derived for each individual correlation coefficient in the moderator analysis using Johnson's DSTAT 1.10 meta-analysis software (1993). When significant moderator effects were found, the proportion of variance in the variables that were accounted for by the moderators was determined. Variables that were significant moderators were entered into a weighted linear multiple regression model (weighted by N – 3; Rosenthal, 1991) using SPSS to clarify their independent effects for explaining variations in z_r. Then, threshold (IQ 120) was added to the previous weighted linear multiple regression model to check the *R*? change to clarify whether the threshold theory had its independent effect for explaining a variation in z_r.

Results

A stem-and-leaf display of the 447 effects is presented in Figure 1, indicating a nearly normal distribution of correlation coefficients. The mean value of *r* was .137 (95% CI = .128 – .146), or .174 (95% CI = .165 – .183) after weighting by sample size. Thus, the mean correlation coefficient was judged to be small and statistically significant. However, the correlation coefficients were heterogeneous, *Q*(446) = 937.058, $p < .0001$; residual standard deviation was greater than one fourth of population correlation coefficient size (.044); and percent of observed variance accounted for by sampling error was 47.64%, which was less than 75%.

Because the data set was heterogeneous, searching for moderators that may account for the remaining systematic variation in the data was necessary (see Schwarzer, 1991). Analyses of moderators were performed by breaking down the data into at least three subsets with respect to theoretically relevant variables. Moderator analyses were conducted to determine whether moderators describing subjects or features of IQ and creativity tests might account for variability in the magnitude of correlation coefficients. Each simple contrast tested as *z* at $p < .001$. Correlation coefficients for each moderator are reported in Tables 1–8 as mean *r*. Variability in the magnitude of correlation coefficients between creativity test scores and IQ scores among different IQ tests (*QB* [6] = 170.193, $p < .0001$), different IQ levels (*QB* [4] = 55.441, $p < .0001$), threshold (*QB*[2] = 17.625, $p < .001$), different creativity tests (*QB*[3] = 203.079, $p < .0001$), creativity test types (*QB* [3] = 34.718, $p < .0001$), creativity subscales (*QB* [4] = 88.380, $p < .0001$), gender (*QB* [2] = 19.163, $p < .0001$), and age (*QB* [4] = 223.282, $p < .0001$) were significant.

Threshold

As Table 1 shows, the mean of the 65 correlation coefficients for below IQ 120 was .201. The correlation coefficients were heterogeneous, $Q(64) = 159.170$, $p > .0001$. The mean of the 14 correlation coefficients for above IQ 120 was .235. The correlation coefficients were heterogeneous, $Q(13) = 45.654$, $p > .0001$. These heterogeneities observed within the different levels or classes imply that the mean correlation coefficients for each level or class cannot be adequately described with a single correlation coefficient. In other words, the variation in the relationships is due to one or more additional factors that the levels or classes do not capture.

Post hoc contrasts revealed no statistically significant differences among below and above the threshold and unreported, likely because there were so few correlation coefficients for above IQ 120. On an a priori basis, IQ below ($r = .201$) and above 120 ($r = .235$) were not statistically different either, $\chi^2 = 2.004$, $p = .157$.

Table 1

Threshold as a Moderator

Threshold	*N* (# of *r*)	Mean *r*	Homogeneity
Above IQ 120	65	.201	heterogeneous
Below IQ 120	14	.235	heterogeneous
Unreported	368.163heterogeneous		

Note. No statistically significant differences between the groups ($p > .001$).
Homogeneity: heterogeneous when $p < .001$.
Model for Threshold, $QB(2) = 17.625$ ($p < .001$).

IQ Levels

As Table 2 shows, the mean of the 32 correlation coefficients for IQ < 100 was .260. The correlation coefficients were homogeneous, $Q(31) = 44.500$, $p = .134$. The mean of the 33 correlation coefficients for 100 < IQ > 120 was .140. The correlation coefficients were heterogeneous, $Q(32) = 98.462$, $p > .0001$. The mean of the 13 correlation coefficients for 120 < IQ > 135 was .259. The correlation coefficients were heterogeneous, $Q(12) = 39.841$, $p < .001$. The mean of the two correlation coefficients for IQ > 135 was -.215. The correlation coefficients were homogeneous, $Q(1) = .234$, $p = .890$.

Post hoc contrasts revealed no statistically significant differences among the different IQ levels. On an a priori basis, however, IQ < 100 ($r = .260$) and 100 < IQ > 120 ($r = .140$) were statistically different, $\chi^2 = 16.208$, $p < .0001$. IQ < 100 ($r = .260$); IQ > 135 ($r = -.215$) were statistically different, $\chi^2 = 12.252$, $p < .001$. 100 < IQ > 120 ($r = .140$); 120 < IQ > 135 ($r = .259$) were statistically different, $\chi^2 = 18.261$, $p < .0001$. 120 < IQ > 135 ($r = .259$); and IQ > 135 ($r = -.215$) were statistically different, $\chi^2 = 12.307$, $p < .001$.

Table 2

IQ Levels as a Moderator

IQ Level	*N* (# of *r*)	Mean *r*	Homogeneity
IQ < 100	32	.260	homogeneous
100 < IQ > 120	33	.140	heterogeneous
120 < IQ > 135	13	.259	heterogeneous
IQ > 135	2	-.215	homogeneous
Unreported	367	.162	heterogeneous

Note. No statistically significant differences between different IQ levels ($p > .001$).
Homogeneity: heterogeneous when $p < .001$.
Model for IQ levels, $QB(4) = 55.441$ (p < .0001).

IQ Tests

As Table 3 shows, the mean of the 21 correlation coefficients for the Terman Concept Mastery Test (TCMT) was .260. The correlation coefficients were homogeneous, $Q(20) = 41.341$, $p > .001$. The mean of the 114 correlation coefficients for the California Test of Mental Maturity (CTMM) was .191. The correlation coefficients were heterogeneous, $Q(113) = 240.652$, $p > .0001$. The mean of the 60 correlation coefficients for the Wechsler Intelligence Scale for Children (WISC) was .097. The correlation coefficients were homogeneous, $Q(59) = 52.377$, $p = .500$. The mean of the 40 correlation coefficients for the School and College Ability Test (SCAT) was .084. The correlation coefficients were homogeneous, $Q(39) = 33.927$, $p = .515$. The mean of the 100 correlation coefficients for the Sequential Tests of Educational Progress (STEP) was .096. The correlation coefficients were homogeneous, $Q(99) = 95.872$, $p = .795$. The mean of the 18 correlation coefficients for the Peabody Picture Vocabulary Test (PPVT) was .018. The correlation coefficients were homogeneous, $Q(17) = 32.922$, $p = .017$.

Post hoc contrasts revealed statistically significant differences ($p < .0001$) between CTMM and STEP. On an a priori basis, CTMM (r = .191) and STEP ($r = .096$) were also statistically different, $\chi^2 = 27.943$, $p < .0001$.

Table 3

IQ Tests as a Moderator

IQ Test	*N* (# of *r*)	Mean *r*	Homogeneity	Contrast	*p*-value for contrast
TCMT	21	.187	homogeneous		
CTMM	114	.191	heterogeneous	CTMM/ STEP	< .0001
WISC	60	.097	homogeneous		
SCAT	40	.084	homogeneous		
STEP	100	.096	homogeneous	CTMM/ STEP	< .0001
PPVT	18	.018	homogeneous		
Others	94	.219	heterogeneous		

Note. TCMT = Terman Concept Mastery Test; CTMM = California Test of Mental Maturity; WISC = Wechsler Intelligence Scale for Children; SCAT = School and College Ability Test; STEP = Sequential Tests of Educational Progress; PPVT = Peabody Picture Vocabulary Test.
Homogeneity: heterogeneous when $p < .001$.
Model for IQ tests, $QB(6) = 170.193$ ($p < .0001$

Table 4

Creativity Tests as a Moderator

Creativity Test	*N* (# of *r*)	Mean *r*	Homogeneity	Contrast	*p*-value for contrast
Guilford	64	.250	heterogeneous	Guilford/Wallach-K	< .0001
TTCT	18	.218	homogeneous		
Wallach-K	319	.116	homogeneous	Guilford/Wallach-K	< .0001
Others	46	.242	heterogeneous		

Note. Guilford = Guilford divergent thinking tasks; TTCT = Torrance Tests of Creative Thinking; Wallach-K = Wallach & Kogan Divergent Thinking Tasks
Homogeneity: heterogeneous when $p < .001$.
Model for Creativity tests, $QB(3) = 203.079$ ($p < .0001$).

Creativity Tests

As Table 4 shows, the mean of the 64 correlation coefficients for the Guilford Tests was .250. The correlation coefficients were heterogeneous, $Q(63) = 175.953$, $p >$ .0001. The mean of the 18 correlation coefficients for the Torrance Tests of Creative Thinking (TTCT) was .218. The correlation coefficients were homogeneous, $Q(17) = 30.252$, $p = .035$. The mean of the 319 correlation coefficients for the Wallach-Kogan divergent thinking measures was .116. The correlation coefficients were homogeneous, $Q(318) = 402.833$, $p > .001$.

Post hoc contrasts revealed statistically significant differences ($p < .0001$) between Guilford Tests and Wallach-Kogan divergent thinking measures. On an a priori basis, Guilford Tests ($r = .250$) and Wallach-Kogan divergent thinking measures ($r = .116$) were also statistically different, $\chi^2 = 145.689$, $p < .0001$.

Creativity Test Types

As Table 5 shows, the mean of the 357 correlation coefficients for verbal tests was .160. The correlation coefficients were heterogeneous, $Q(356) = 738.744$, $p > .0001$. The mean of the 41 correlation coefficients for non verbal tests was .226. The correlation coefficients were heterogeneous, $Q(40) = 84.713$, $p < .0001$. The mean of the 46 correlation coefficients for mixed tests was -.235. The correlation coefficients were homogeneous, $Q(45) = 77.314$, $p > .001$.

Post hoc contrasts revealed no statistically significant differences ($p > .001$) between verbal tests and nonverbal tests. On an a priori basis, however, verbal tests ($r = .160$) and nonverbal tests ($r = .226$) were statistically different, $\chi^2 = 11.522$, $p < .001$.

Table 5

Types of Creativity Tests as a Moderator

Creativity Test Type	N (# of r)	Mean r	Homogeneity
Verbal	357	.160	heterogeneous
Nonverbal	41	.226	heterogeneous
Mixed	46	.235	homogeneous
Unreported	3	.068	homogeneous

Note. No statistically significant differences between different Creativity Test Types ($p > .001$).
Homogeneity: heterogeneous when $p < .001$.
Model for Creativity Test Types, $QB(3) = 34.718$ ($p < .0001$).

Creativity Subscales

As Table 6 shows, the mean of the 175 correlation coefficients for originality was .131. The correlation coefficients were heterogeneous, $Q(174) = 267.721$, $p > .0001$. The mean of the 184 correlation coefficients for fluency was .170. The correlation coefficients were heterogeneous, $Q(183) = 370.085$, $p > .0001$. The mean of the six correlation coefficients for figural redefinition was .362. The correlation coefficients were homogeneous, $Q(5) = 2.531$, $p = .865$. The mean of the 25 correlation coefficients for flexibility was .231. The correlation coefficients were heterogeneous, $Q(24) = 82.439$, $p > .0001$.

Post hoc contrasts revealed statistically significant differences ($p < .0001$) between originality and figural redefinition, between originality and flexibility, and between fluency and figural redefinition. On an a priori basis, originality ($r = .131$) and fluency ($r = .170$) were statistically different, $\chi^2 = 13.241$, $p < .001$. Originality ($r = .131$) and figural redefinition ($r = .362$) were statistically different, $\chi^2 = 47.177$, $p < .0001$. Originality ($r = .131$) and flexibility ($r = .231$) were statistically different, $\chi^2 = 38.477$, $p < .0001$. Fluency ($r = .170$) and figural redefinition ($r = .362$) were statistically different, $\chi^2 = 33.766$, $p < .0001$. Fluency ($r = .170$) and flexibility ($r = .231$) were statistically different, $\chi^2 = 15.143$, $p < .0001$. Figural redefinition ($r = .362$) and flexibility ($r = .231$) were statistically different, $\chi^2 = 14.624$, $p < .001$.

Gender

As Table 7 shows, the mean of the 186 correlation coefficients for males was .149. The correlation coefficients were heterogeneous, $Q(185) = 422.221$, $p > .0001$. The mean of the 180 correlation coefficients for females was .159. The correlation coefficients were homogeneous, $Q(179) = 213.439$, $p = .087$.

Post hoc contrasts revealed no statistically significant differences ($p = .736$) between males and females. On an a priori basis, males ($r = .149$) and females ($r = .159$) were not statistically different either, $\chi^2 = .614$, $p = .433$.

Age

As Table 8 shows, the mean of the 251 correlation coefficients for the elementary school group was .086. The correlation coefficients were homogeneous, $Q(250) = 260.451$, $p = .660$. The mean of the 27 correlation coefficients for the middle school group was .210. The correlation coefficients were heterogeneous, $Q(26) = 70.097$, $p > .0001$. The mean of the 105 correlation coefficients for the high school group was .261. The correlation coefficients were heterogeneous, $Q(104) = 163.761$, $p > .001$. The mean of the 53 correlation coefficients for adults was .205. The correlation coefficients were heterogeneous, $Q(52) = 183.404$, $p > .0001$.

Post hoc contrasts revealed statistically significant differences ($p < .0001$) between the elementary school group and the middle school group, between the elementary school group and the high school group, and between the elementary school group and adults. On an a priori basis, the elementary school group ($r = .086$) and the middle school group ($r = .210$) were statistically different, $\chi^2 = 59.051$, $p < .0001$. The elementary school group ($r = .086$) and the high school group ($r = .261$) were statistically different, $\chi^2 = 110.148$, $p < .0001$. The elementary school group ($r = .086$) and adults ($r = .205$) were statistically different, $\chi^2 = 126.945$, $p < .0001$. The high school group ($r = .261$) and adults ($r = .205$) were statistically different, $\chi^2 = 12.036$, $p < .001$.

Multiple Regression Analysis

Variables that were statistically significant moderators (based on the results of post hoc contrasts) were entered into a weighted linear multiple regression model to determine their independent effects for explaining variation in the magnitude of correlation coefficients. The results of the direct entry of the four statistically significant moderating variables of IQ tests, creativity tests, creativity subscales, and age into a weighted multiple linear regression analysis indicated that age and creativity subscales independently accounted for variation in z_r. The regression model yielded $R? = .260$, adjusted $R? = .235$, $F(221) = 10.716$, $p < .0001$. Next, the threshold variable was added to the previous weighted multiple linear regression analysis

Table 6

Creativity Subscales as a Moderator

Creativity Subscale	*N* (# of *r*)	Mean *r*	Homogeneity	Contrast	*p*-value for contrast
Originality	175	.131	heterogeneous	ac, ad	< .0001
Fluency	184	.170	heterogeneous	bc	< .0001
Figural Redefinition	6	.362	homogeneous	ac, bc	< .0001
Flexibility	25	.231	heterogeneous	ad	< .0001
General Creativity	57	.206	heterogeneous		

Note. Homogeneity: heterogeneous when $p < .001$.
Model for Creativity Subscales, $QB(4) = 88.380$ ($p < .0001$).

model. Adding the threshold variable did not change the *R*? significantly. The regression model yielded *R*? Change = .004, *F* Change (1) = 1.269, $p = .261$.

According to Johnson (1993), the tests of meta-analytic predictor's significance are inappropriate because the error degrees of freedom in the weighted regression procedure of SPSS are too large. Thus, the appropriate corrections in its regression model were made using DSTAT. The results of the weighted multiple linear regression of effect size z_r on moderator variables are presented in Table 9. The various creativity tests ($\beta = -.01179$, $z = -2.33020$, $p < .001$) and age ($\beta = .03917$, $z = 6.19332$, $p < .0001$) had significant effects on the magnitude of the correlation coefficients between creativity test scores and IQ test scores. Various IQ tests ($\beta = -.00170$, $z = -.53790$, $p = .59.64$) and creativity subscales ($\beta = .01647$, $z = 3.72019$, $p = .01980$) were uncorrelated with the magnitude of the correlation coefficients. The results of the weighted multiple linear regression of effect size z_r with the threshold variable are presented in Table 10. The threshold variable ($\beta = -.00726$, $z = -.76628$, $p = .44351$) was not related to the magnitude of the correlation coefficients.

Table 7

Gender as a Moderator

Gender Group	*N* (# of *r*)	Mean *r*	Homogeneity
Male	186	.149	heterogeneous
Female	180	.159	homogeneous
Combined	81	.193	heterogeneous

Note. No statistically significant differences between male and female groups ($p > .001$).
Homogeneity: heterogeneous when $p < .001$.
Model for Gender, $QB(2) = 19.163$ ($p < .0001$).

Discussion

The quantitative synthesis of the literature results indicate that the relationship between IQ scores and creativity scores is small and positive, in which the mean effect *r* was .174 (95% CI, .165 - .183). The correlation coefficients were heterogeneous, but the threshold of IQ 120, which was examined as one of the possible moderators, could not explain variance in the studies' correlation coefficients. Further, when the IQ scores were divided into the four levels (IQ < 100 [$r = .260$]; 100 < IQ > 120 [$r = .140$]; 120 < IQ > 135 [$r = .259$]; IQ > 135 [$r = -.215$]), there were no statistically significant differences among the levels according to the results of the post hoc contrasts, although IQ > 135 had a negative mean correlation coefficient, $r = -.215$. Thus, the threshold theory was not supported.

The results of the post hoc contrasts revealed that IQ tests, creativity tests, creativity subscales, and age explained the differences in correlation coefficients between IQ and creativity tests scores. However, the variance in the magnitude of the correlation coefficients was not significantly explained by the various IQ tests, subscales of the creativity tests, creativity test types, and gender, but it was significantly explained by different creativity tests ($p < .001$) and age ($p < .0001$) according to the results of the weighted multiple linear regression, which determines moderators' independent effects for explaining variation.

Post hoc contrasts revealed a statistically significant difference between Guilford Tests and Wallach-Kogan divergent thinking measures. According to the results of the weighted multiple linear regression, various creativity tests had significant independent effects on the magnitude of the correlation coefficients between creativity test

Table 8

Age as a Moderator

Age Group	*N* (# of *r*)	Mean *r*	Homogeneity	Contrast	*p*-value for contrast
Elementary	251	.086	homogeneous	ab, ac, ad	< .0001
Middle	27	.210	heterogeneous	ab	< .0001
High	105	.261	heterogeneous	ac	< .0001
Adult	53	.205	heterogeneous	ad	< .0001
Unreported	11	.267	heterogeneous		

Note. Homogeneity: heterogeneous when $p < .001$.
Model for Age, $QB(4) = 223.282$ ($p < .0001$).

Table 9

Multiple Linear Regression of Effect Size z_r With Threshold (Weighted by Sample Size)

Moderator	Standardized	*z*-value	*p*-value
IQ Tests	-.00170	-.53790	.59064
Creativity Tests	-.01179	-2.33020	< .001
Creativity Subscales	.01647	3.72019	.01980
Age	.03917	6.19332	< .0001

Note. Overall regression effect = 48.96, $df = 4$, $p < .0001$ (two-tailed).

Table 10

Multiple Linear Regression of Effect Size z_r on Moderator Variables (Weighted by Sample Size)

Moderator	Standardized	*z*-value	*p*-value
IQ Tests	.00085	-.26862	.78822
Creativity Tests	-.01143	-2.26266	< .001
Creativity Subscales	.01638	3.70578	.02366
Age	.03976	6.29666	< .0001
Threshold	-.00726	-.76628	.44351

Note. Overall regression effect = 49.565, $df = 5$, $p < .0001$ (two-tailed).

scores and IQ test scores. The mean weighted correlation coefficients between IQ test scores and Wallach-Kogan divergent thinking measure scores, $r = .116$, was much smaller than the mean weighted correlation coefficients between IQ tests scores and Guilford Test scores, $r = .250$. This might be because Wallach-Kogan divergent thinking measures were administered as a game-like activity, whereas the Guilford Tests were administered as tests.

Wallach and Kogan (1965) concluded that traditional intelligence test scores are independent from creativity test scores when the creativity tests are assessed in a game-like, nonevaluative context. Some other studies also support this conclusion (e.g., Cropley & Maslany 1969; Iscoe & Pierce-Jones, 1964; Kogan & Pankove, 1972; Pankove & Kogan, 1968; Wallach & Wing, 1969; Ward, 1968). Torrance (1966) also recommended the creation of a game-like, thinking, or problem-solving atmosphere in order to avoid the threatening situation associated with testing. His intent was to set a tone so that the examinees would expect to enjoy the activities.

Examinees should be encouraged to "have fun," and they should experience a psychological climate that is as comfortable and stimulating as possible (see Kim, in press). Thus, according to the TTCT manual (Ball & Torrance, 1984), administrators of the tests should invite the examinees to enjoy the activities and view the tests as a series of fun activities, thereby reducing test anxiety. Torrance (1987) suggested a warm-up activity before administration of the TTCT for arousing the incubation processes and increasing motivation, thereby enabling creative energy. Despite Torrance's recommendation, however, none of the studies using the TTCT included in the present study reported that the TTCT was administered in a game-like testing context. Thus, how many administrators of the TTCT included in this study invited the examinees to enjoy the activities cannot be known. That might be why the mean correlation coefficient of the TTCT was .218, which was not significantly different from either test-like Guilford Tests ($r = .250$) or game-like

Wallach-Kogan divergent thinking measures (r = .116). Therefore, it can be concluded that, when creativity tests are administered in a game-like testing context, the creativity test scores have smaller relationships with IQ test scores.

The variance in the magnitude of the correlation coefficients was also significantly explained by the variance among age groups. For older groups (middle school, high school, and adult groups), IQ scores were more closely associated with creativity scores than within the younger group (kindergarten through fifth-grader group). Given that even the same tests may measure different processes when administered to subjects of differing ages (Ward, 1968), creativity tests may also measure different constructs among various ages. This is supported by previous studies (Kim, 2004; Kim, Cramond, & Bandalos, 2004, in press) that show that the latent structure of the TTCT is more invariant across gender than across grade groups. The relationship between creativity and IQ scores among younger children might be smaller because of less educational influence over the use of their cognitive abilities, as compared with older people.

In conclusion, the negligible relationship between creativity and IQ scores indicates that even students with low IQ scores can be creative. Therefore, teachers should be aware of characteristics of creative students—this will enable teachers to see the potential of each child. In contrast with the threshold theory, neither IQ 120 nor different levels of IQ scores explained variation in the correlation coefficients. The differences in the correlation coefficients between IQ scores and creativity test scores were not significantly explained by the various IQ tests, subscales of the creativity tests, or creativity test types, whereas different creativity tests and age were consistent moderator variables. The significantly independent moderator effect for creativity tests indicates that, when creativity tests are administered in a game-like context, the creativity test scores have smaller relationships with IQ test scores than when creativity tests are administered in a test-like context. The significantly independent moderator effect for age, in which IQ scores were more closely associated with creativity scores for younger groups than for older groups, indicates less educational influence on students' use of their cognitive abilities for younger groups compared with older groups.

However, the findings of the present study in terms of the threshold theory are limited in generalizability because, for the 368 of the 447 correlation coefficients, the subjects' IQ scores were unreported or unable to be used. For future studies, more studies that have reported IQ scores are needed.

References

Ball, O. E., & Torrance, E. P. (1984). *Streamlines scoring workbook: Figural A*. Bensenville, IL: Scholastic Testing Service.

Barron, F. (1961). Creative vision and expression in writing and painting. In D. W. MacKinnon (Ed.), *The creative person* (pp. 237–251). Berkeley: Institute of Personality Assessment Research, University of California.

Cooper, H., & Hedges, L. V. (1994). Research synthesis as a scientific enterprise. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 3–14). New York: Russell Sage Foundation.

Cropley, A. J., & Maslany, G. W. (1969). Reliability and factorial validity of the Wallach-Kogan creativity tests. *British Journal of Psychology, 60*, 395–398.

Flescher, I. (1963). Anxiety and achievement of intellectually gifted and creatively gifted children. *Journal of Psychology, 56*, 251–268.

Getzels, J. W., & Jackson, P. W. (1958). The meaning of "giftedness": An examination of an expanding concept. *Phi Delta Kappan, 40*, 75–77.

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

Gough, H. G. (1976). Studying creativity by means of word association tests. *Journal of Applied Psychology, 61*(3), 348–353.

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

Guilford, J. P. (1962). Potentiality for creativity and its measurement. In *Proceedings of the 1962 invitational conference on testing problems* (pp. 31–39). Princeton, NJ: Educational Testing Service.

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

Guilford, J. P., & Christensen, P. R. (1973). The one-way relation between creative potential and IQ. *Journal of Creative Behavior, 7*, 247–252.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.

Helson, R., & Crutchfield, R. S. (1970). Mathematicians: The creative researcher and the average Ph.D. *Journal of Consulting and Clinical Psychology, 34*, 250–257.

Helson, R., & Crutchfield, R. S. (1971). Women mathematicians and the creative personality. *Creative Achievements, 32*, 210–220.

Herr, E. L., Moore, G. D., & Hansen, J. C. (1965). Creativity, intelligence, and values: A study of relationships. *Exceptional Children, 32*, 114–115.

Hocevar, D. (1980). Intelligence, divergent thinking, and creativity. *Intelligence, 4,* 25–40.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies.* Beverly Hills, CA: Sage.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.

Iscoe, I., & Pierce-Jones, J. (1964). Divergent thinking, age, and intelligence in White and Negro children. *Child Development, 35*, 785–798.

Johnson, B. T. (1993). *DSTAT 1.10: Software for the meta-analytic review of research literatures.* Hillsdale, NJ: Erlbaum.

Kim, K. H. (2004, April 30). *Confirmatory factor analyses and multiple group analyses on the Torrance Tests of Creative Thinking*. Poster session presented at Research Symposium: The Preparation of Educators, University of Georgia, Athens.

Kim, K. H. (in press). Can we trust creativity tests?: A review of The Torrance Tests of Creative Thinking (TTCT). *Creativity Research Journal.*

Kim, K. H., Cramond, B., & Bandalos, D. L. (in press). The latent structure and measurement invariance of scores on the Torrance Tests of Creative Thinking–Figural. *Educational and Psychological Measurement.*

Kogan, N., & Pankove, E. (1972). Creative ability over a five-year span. *Child Development, 43,* 427–442.

MacKinnon, D. W. (1961). Creativity in architects. In D. W. MacKinnon (Ed.), *The creative person* (pp. 291–320). Berkeley: Institute of Personality Assessment Research, University of California.

MacKinnon, D. W. (1962). The nature and nurture of creative talent. *American Psychologist, 17*, 484–495.

MacKinnon, D. W. (1967). Educating for creativity: A modern myth? In P. Heist (Ed.), *Education for creativity* (pp. 1–20). Berkeley, CA: Center for Research and Development in Higher Education.

Pankove, E., & Kogan, N. (1968). Creative ability and risk taking in elementary school children. *Journal of Personality, 36*, 420–439.

Richards, R. L. (1976). A comparison of selected Guilford and Wallach-Kogan creative thinking tests in conjunction with measures of intelligence. *Journal of Creative Behavior, 10*(3), 151–164.

Rosenthal, R. (1991). *Meta-analytic procedures for social research.* Beverly Hills, CA: Sage.

Rosenthal, R., & Rubin, D. B. (1984). Multiple contrasts and ordered Bonferroni procedures. *Journal of Educational Psychology, 76*, 1028–1034.

Rossman, B. B., & Horn, J. L. (1972). Cognitive, motivational and temperamental indicants of creativity and intelligence. *Journal of Educational Measurement, 9*, 265–286.

Rotter, D. M., Langland, L., & Berger, D. (1971). The validity of tests of creative thinking in seven-year-old children. G*ifted Child Quarterly, 15*, 273–278.

Runco, M. A., & Albert, R. S. (1986). The threshold theory regarding creativity and intelligence: An empirical test with gifted and nongifted children. C*reative Child and Adult Quarterly, 11*, 212–218.

Schwarzer, R. (1991). *Meta: Programs for secondary data analysis 5.3*. Berlin: Free University of Berlin.

Simonton, D. K. (1994). *Greatness: Who makes history and why.* New York: Guilford Press.

Torrance, E. P. (1977). *Creativity in the classroom.* Washington, DC: National Education Association.

Torrance, E. P. (1987). *Guidelines for administration and scoring/comments on using the Torrance Tests of Creative Thinking*. Bensenville, IL: Scholastic Testing Service.

Vernon, P. E. (1972). The validity of divergent thinking tests. Walberg, H. A. (1988). Creativity and talent as learning. In R. J. Sternberg (Ed.),