Do Gifted Student Writers and Creative Writing Experts Rate Creativity the Same Way?

James C. Kaufman  
Learning Research Institute  
California State University at San Bernardino

Claudia A. Gentile  
Educational Testing Service

John Baer  
Rider University

ABSTRACT

Little research has been conducted on how gifted novices compare to experts in their judgments of creative writing. If novices and experts assign similar ratings, it could be argued that gifted novices are able to offer their peers feedback of a similar quality to that provided by experts. Such a finding would support the use of collaborative feedback in gifted classrooms. We asked gifted high school creative writers and three groups of experts (cognitive psychologists, creative writers, and teachers) to rate a set of 27 short stories and 28 poems for creativity using a scale of 1 to 6. The interrater agreement among the novices was within acceptable standards, and the agreement among the experts was very strong. When the ratings of novices were compared to the ratings of experts, a strong degree of correlation was found, supporting the use of peer feedback among gifted novice creative writers.

How do we “grade” a creative work? How can we measure the creativity of a student’s poem or short story? One suggestion from the field of creativity research is the consensual assessment technique (Amabile, 1982, 1996). In this technique, expert raters, selected or recruited for their experience in a domain, use their own ideas about creativity to provide independent ratings of a product. These judgments of creativity are given without consulting with either the experimenter or fellow raters.

Yet, what does it take to be an “expert” rater? Can nonexpert raters reach consensus and provide appropriate judgments of creativity? How do expert and novice raters compare? This issue has been explored for nearly a century under the name “aesthetic judgment” (Cattell, Glascott, & Washburn, 1918). Many past investigations have found that expert-level judges consistently agree and have high interjudge reliabilities when judging artistic works (e.g., Child, 1962), even across different cultures (Child & Iwao, 1968; Haritos-Fatouros & Child, 1977; Iwao, Child, & Garcia, 1969; Rostan, Pariser, & Gruber, 2002). Some initial work has been conducted on comparing novice and expert judgments in domains such as cognitive science or creativity (Amabile, 1996: Baer, 1997b). As such, teachers and others who wish to help their students write more creatively must be especially careful about the kinds of evaluations they provide their students (Amabile, 1989; Baer, 1997a).

Putting the Research to Use

Creative writing students need feedback to improve their writing, but not all feedback is equally helpful. Evaluation always has the potential to influence creativity negatively because it tends to increase extrinsic motivation and decrease intrinsic motivation (Amabile, 1996; Baer, 1997b). As such, teachers and others who wish to help their students write more creatively must be especially careful about the kinds of evaluations they provide their students.

It perhaps goes without saying that feedback from experts will generally be more valuable than feedback from nonexperts. But, who qualifies as an expert? Teachers sometimes have students read and respond to other students’ writing. Unfortunately, this runs the risk of exposing student writers to the risks of evaluation without any assurance that the feedback they are getting is likely to be helpful.

This study provides evidence that, when it comes to overall evaluations of the creativity of student writing, gifted student writers are a decent substitute for the kinds of experts we would like to have evaluating student work. This doesn’t guarantee that every gifted student writer will be a good judge, of course, any more than we should expect every judgment by a recognized expert (e.g., a writer, writing teacher, or literary critic) to be helpful. But, it does increase our confidence that peer feedback will be useful and productive in a class of gifted student writers.
artwork (Haritos-Fatouros & Child; Runco, McCarthy, & Svenson, 1994), dramatic performance (Myford, 1989), and music (Hickey, 2001).

This question has not been adequately addressed for the issue of novice versus expert judgments of creative writing, however. Amabile (1982, 1996) has examined the reliabilities of different judges, including teachers, poets, and graduate students in creative writing, for haiku poems and stories. She has found high levels of reliability for all groups (ranging from .77 to .91), and her consensus assessment technique has been widely used in creativity research with generally high levels of interrater reliability (e.g., Baer, 1993, 1994, 1997b, 1998; Baer, Kaufman, & Gentle, 2004; Hennessey & Amabile, 1999; Runco, 1989). However, no research has been conducted comparing the ratings of experts in the domain of creative writing with those of novice creative writers. In addition, little work has focused on gifted novices and experts.

The lack of work on the appropriateness of novice judges rating a piece of creative writing is surprising given the focus in schools on collaborative learning and peer conferencing. Students are often encouraged to learn not only from the teacher, but also from each other (Alber, 1999; Burke, 1994; Clark, 1983; Elbow, 1973; Elbow & Belanoff, 1999; Gorman, 1998; Graham & Harris, 1988; Jeske, 1989; Karegianes, Pascarella, & Pflaum, 1980; MacArthur, Graham, Schwartz, & Schafer, 1995; Mayfair, 1999; Mueller & Fleming, 2001; Salend, 1990). Such evaluation of writing by peers is said to provide both “ample feedback on their drafts, which enables them to do a thoughtful, informed revision [and] valuable editing practice, which enables them to edit their own work better in the future” (Jeske, 2002). The advent of computers and the Internet has allowed students to interact with a much larger peer group than ever before (McFadzean & McKenzie, 2001).

Gifted students are particularly good candidates for judging the merits of a piece of creative writing (Clark, 1983). They tend to excel at metacognition, the ability to monitor one’s own learning, perform self-evaluation, and then make plans accordingly (see Everson & Tobias, 1998; Flavell, 1979). In particular, gifted students are better able to transfer strategies into different contexts (Robinson & Clinkenbeard, 1998). A gifted student who learned about the components of quality creative writing might be more likely to transfer these ideas into evaluating a peer’s creative writing.

When students are asked to evaluate the quality of their peers’ writing in collaborative learning situations, however, some critics claim that this kind of learning activity is not that helpful because students’ judgments are those of novice writers (Jeske, 1989). Even if these students are gifted and more likely to make more mature and insightful comments, these comments may still not be comparable to the quality of feedback an expert adult (such as a teacher) would offer a student writer.

Yet, little research has been conducted on how gifted novices compare to experts in their judgments of creative writing. If novices and experts assign highly correlated ratings to pieces of creative writing, then it could be argued that gifted novices are able to offer their peers feedback of a similar quality to that which experts could provide. Such a finding would support the use of collaborative feedback in gifted classrooms.

To investigate this question, we asked a team of high school creative writers attending the New Jersey Governor’s School of the Arts to rate a set of 27 short stories and 28 poems for creativity using a scale of 1 to 6. We also asked three groups of experts to rate the same creative pieces on the same scale so that the consistency in ratings between the novices and experts could be investigated.

**Method**

**Selection of Materials**

The 27 short stories and 28 poems were drawn from the 1998 National Assessment of Educational Progress (NAEP) Classroom Writing Study. In that study, eighth graders from 32 states were asked to assemble folders containing two samples of their best writing. Approximately 125 classrooms, representing a wide variety of demographics, participated in this study. For the present study, a subsample of 27 short stories and 28 poems was selected for analysis. The papers selected represented a range of community types (rural, suburban, urban) and major geographic regions of the country (Northeast, Southeast, Central, and West). No more than one paper per student was included in the samples. Although an attempt was made to select papers from multiple regions and communities, selection was not based on content or quality of the papers.

**Procedure**

Eight gifted creative writers were recruited from among the high school juniors selected to attend the New
Jersey Governor’s School of the Arts. From the many students who apply to the creative writing program of the Governor’s School of the Arts, only 10% are accepted. This rigorous process involves an evaluation of student writing samples and a subsequent interview. Thus, the participants in this study represent some of the most gifted “novice” creative writers in the state.

In addition, 13 expert judges participated in this study, representing three different types of expertise. The first type of expertise involved an in-depth familiarity with eighth-grade creative writing. Middle school teachers who emphasized the teaching of creative writing in their practice comprised this first group of experts. The second group of experts consisted of published creative writers, all of whom also had extensive experience working with middle school students, either through giving workshops in the schools or through editing collections of creative writing by middle school students. The third type of expertise involved a familiarity with research on creativity. Psychologists who study creativity made up the third group of experts. There was roughly equal representation in each of these types of expertise (four teachers, four writers, and five psychologists). We recruited raters based on three factors: an appropriate level of expertise (e.g., the psychologists had all published papers in the area of creativity); a willingness to undertake the extensive rating procedure; and a proximity to Princeton, NJ, for any possible future work.

Novices and experts rated the poems and short stories independently. They were asked to read the poems and short stories and assign them a score from 1 to 6, with 1 being the lowest level of creativity and 6 representing the highest level of creativity. Raters were asked to assign creativity ratings based on their own personal definition of creativity; no additional guidance, descriptors, or material on creativity was provided.

To help them with the task, judges were encouraged first to divide the papers in each group into three piles (low, medium, and high creativity) and then to subdivide each pile to create six levels of creativity. In their final ratings, they were free to move papers into whichever of the six levels they deemed most appropriate, regardless of their initial rankings, and they were asked to report only their final ratings.

Each rater read every piece of writing. Ratings were conducted and collected entirely through the mail. Raters did not meet or talk about their ratings with one another or with the experimenters until after all the ratings had been submitted.

Data Analysis

For this analysis, we were interested in two questions. The first question was “How much did raters in each group (novice and expert) agree with each other?” For this question, we conducted coefficient alpha interrater reliability analysis, a recommended technique (Fleiss, 1981; Landis & Koch, 1977). The second question was “What is the relationship of the novice raters to the expert raters?” For this question, we computed Pearson correlation coefficients.

Results

In order to address the question of whether gifted novice ratings are similar to expert ratings, we first investigated interrater reliabilities among the novice group and among the expert group. Then, we compared the novice ratings with the combined expert ratings and also with the ratings of different subgroups of experts.

For the novices, the coefficient alpha interrater reliabilities were .82 for the poems and .74 for the short stories. For the experts, the coefficient alpha interrater reliabilities were .88 for the poems and .88 for the short stories. Interrater correlation coefficients in the area of .75 are considered to be “excellent” (Fleiss, 1981). Indeed, Landis and Koch (1977) posited that any interrater correlation above .80 is “almost perfect.” Thus, the interrater agreement among the novices was within acceptable standards, and the agreement among the experts was very strong.

When the ratings of novices were compared to the ratings of experts, a strong degree of correlation was found. Full correlations are presented in Table 1. Novices’ ratings correlated with expert raters at \( r = .78 \) for poetry and \( r = .77 \) for short stories. As can be seen in Table 1, the correlations for the three different types of experts across poetry and short stories ranged from .62 to .80. Novice ratings of poetry and short stories both correlated highest with the expert group of writers. All correlations were significant at \( p < .0001 \).

Discussion

Gifted novices produced ratings with nearly as high a reliability as did experts. In addition, the gifted novices’ ratings correlated significantly with all three expert groups. This finding indicates that gifted novices may
Table 1
Correlations Between Novice and Different Expert Raters by Type of Writing

<table>
<thead>
<tr>
<th></th>
<th>Poetry</th>
<th>Short Stories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychologists</td>
<td>.62*</td>
<td>.65*</td>
</tr>
<tr>
<td>Writers</td>
<td>.80*</td>
<td>.74*</td>
</tr>
<tr>
<td>Teachers</td>
<td>.68*</td>
<td>.62*</td>
</tr>
<tr>
<td>All Experts</td>
<td>.78*</td>
<td>.77*</td>
</tr>
</tbody>
</table>

Note. * p < .0001

well produce equally effective responses to creative work as experts. This research has implications in two contexts: One is more relevant to creativity and giftedness research and the other is more “hands on” for the classroom.

The distinction between novice and experts in the creativity and giftedness field is one that has been much discussed. On the one hand, some theorists claim that possessing too much knowledge about a domain can prevent truly novel and original thoughts. Frensch and Sternberg (1989), for example, found that expert bridge players found it more difficult to adjust to changes in the rules of the game than novices (both surface changes and conceptual changes, but more notably the latter). Minsky (1997) argued that a great deal of our knowledge is geared toward avoiding negative experiences—and yet it is these very negative experiences that may result in creative production. More specific to this study, Runco, McCarthy, and Svenson (1994) suggested that professionals in an artistic field may be less capable of assessing student work than peers or teachers.

In contrast, other theorists attribute a large portion of creative success to knowledge and expertise. Perhaps most extremely, Ericsson and Charness (1994) argued that experience and extended practice account for much of what distinguishes elite performers. Certainly, there is a great deal of evidence in support of a “10-year” rule—that a creative person’s first significant contribution tends to occur approximately 10 years after first entering a field (Hayes, 1981, 1989).

Gifted novices represent an interesting phenomenon. They have more experience and are farther along into their “10 years.” Yet, they may also be early enough in a career that they may avoid some of the pitfalls that ensnare experts. This study suggests that gifted novices may be close enough to experts that their ability to judge creative work may have merit in its own right. It is important to distinguish, however, the ability to give consistent, appropriate ratings of creative work from the ability to give useful feedback. It may well be the case that gifted novices can recognize the quality of creative work, but are less equipped to articulate why and how individual variations occur.

Another way of interpreting these findings is to examine them in the context of collaborative learning. In an extensive survey of writing education, the National Assessment of Educational Progress asked a nationally representative sample of more than 20,000 eighth graders and more than 14,000 teachers to comment on the use of peer review and discussion in writing instruction (Greenwald, Persky, Campbell, & Mazzeo, 1999). Thirty-three percent of the teachers reported that they always had students discuss their writing with peers, and 64% said that they sometimes did. When students were asked a similar question (“How often does your teacher ask you to work in pairs or small groups to discuss your writing?”), 11% said “almost every day,” 28% said “1–2 times per week,” 33% said “1–2 times per month,” and 28% said “never or hardly ever.” Thus, peer review is a common practice in about one third and a frequent practice in at least another third of the eighth-grade English classrooms.

While peer review of student work is often recommended as a way to improve student writing, some have questioned the quality of the feedback that the novice writers could be expected to provide one another. This study provides support for the use of peer feedback in classes of gifted writers because it demonstrates that, at least in the area of summary judgments of the creativity of a poem or story, gifted creative writers and experts in the field of creative writing tend to give very similar evaluations. Similarly, it suggests that gifted novice creative writers may be able to provide useful feedback—similar to that which experts might provide—to younger or less gifted creative writers.

It is interesting that, among the three subgroups of experts (cognitive psychologists, teachers, and creative writers), the correlations of gifted high school novice creative writers were highest with adult expert creative writers in both the poetry and short story samples. This is not surprising because this is the group into which the novices hope to enter themselves as they and their writing mature.

While our results provide support for the use of peer conferencing and other kinds of peer feedback in classes of gifted creative writers and for the use of gifted
experts to evaluate the work of less gifted writers and provide them feedback, it must be emphasized that the novices in this study were all gifted novices, and any extrapolation of these results to more typical novice writers must be made cautiously. We wish specifically to note that this study should not be used as a carte blanche endorsement of all uses of peer feedback in writing classes. Peer conferencing is used (and widely recommended) among virtually all levels of student writers, even in special education classrooms where writers may exhibit minimal writing talent or interest in writing (Alber, 1999; Graham & Harris, 1988; Karegianes et al., 1980; MacArthur et al., 1995; Salend, 1990). While it may be true that the use of peer evaluations of student writing are useful in such classrooms, this study included only gifted novices and cannot determine whether less gifted novices could provide similarly valuable feedback.

It should also be borne in mind that this study compared ratings of the creativity of short stories and poems of expert judges and gifted novices, not judgments of other aspects of those stories and poems (such as grammatical correctness). While we think it likely that gifted novice writers would also be able to provide accurate and useful feedback in such areas (because it is unlikely they would have been identified as gifted writers without at least a modest level of knowledge of grammar, punctuation rules, etc.), that is not what this study asked them to assess. When peer conferencing is used in regular (i.e., noncreative writing) classrooms, a major focus of the feedback peers would be asked to provide could be in areas such as grammar, punctuation, and capitalization. We strongly caution that using this study—which did not examine any of those areas and used only gifted novice writers—to support the use of peer conferencing in regular, nongifted, noncreative writing classes would be an extreme extrapolation of our data. There is a need for further research in this area to investigate how the evaluations (of both creativity and of other aspects of writing) of nongifted novice writers compare to the evaluations of experts.

In summary, by showing that gifted novices rate the creativity of short stories and poems in a manner similar to that of experts, this study offers support for the quality of gifted novices’ aesthetic judgment. In addition, this study supports the use of peer feedback as used in many collaborative learning situations—at least among gifted students who are on their way to becoming “experts.”

References


**Author Note**

The authors would like to thank Kathy Howell, Susan Martin, Venus Mifusid, and Alyson Tregidgo for their assistance, and Fred Cline for help with data analysis. A special thanks goes to Mike Rothstein, director of the New Jersey Governor’s School of the Arts, for his vital support to the project.

This paper was supported by a grant from the National Center for Educational Statistics.

Please address all correspondence to James C. Kaufman, Learning Research Institute, California State University at San Bernardino, Department of Psychology, 5500 University Parkway, San Bernardino, CA 92407.
About the Authors

John Baer is a professor of educational psychology at Rider University. In 1993, he won the Daniel E. Berlyne Award of the American Psychological Association for his creativity research, and in 1997, the Eighth National Conference on College Teaching and Learning presented him their Award for Innovative Excellence in Teaching, Learning, and Technology. He has published two books about creativity, coedited two forthcoming books on creativity, and published numerous research articles in publications such as the Journal of Creative Behavior and the Creativity Research Journal.

David W. Chan is professor in the Department of Educational Psychology and program supervisor of the Programs for the Gifted and Talented at the Chinese University of Hong Kong. His research interests include stress and coping, psychopathology and health, creativity and intelligence, and gifted education and talent development. E-mail: davidchan@cuhk.edu.hk

Christine F. Delgado is a research assistant professor of psychology and an adjunct assistant professor of education and psychological studies at the University of Miami. She is the co-principal investigator and project director for the Children’s Registry and Information System (CHRIS) project. The CHRIS project develops and maintains the State of Florida database for tracking preschool children with disabilities served under Part B of the Individuals With Disabilities Education Act. Her research focuses on the early identification of children at risk for developmental disabilities, as well as the longitudinal tracking of these children to monitor their educational progress. E-mail: cdelgado@umiami.edu

Joanne F. Foster acquired her master’s degree in special education and adaptive instruction and her doctorate in human development and applied psychology at the Ontario Institute for Studies in Education at the University of Toronto (OISE/UT). She is currently a lecturer of educational psychology in the Preservice Program (high-ability cohort) at OISE/UT and the gifted education resource consultant for a large school board in Toronto. With Dona Matthews, Joanne has cowritten a comprehensive resource book for parents and teachers entitled Being Smart: Understanding and Encouraging Giftedness in Children (in press; for more information, see http://www.beingsmart.ca). She conducts professional learning programs on giftedness, is actively involved with a number of educational committees throughout the community, and speaks to parent forums and educators regionally and internationally on the complexities and implications of high-level development.

Claudia A. Gentile is a literacy educator who specializes in developing innovative approaches to instruction and assessment. She holds a Ph.D. in English education from Syracuse University and has taught courses in linguistics, writing, literacy, and research methods in the U.S. and Japan. Through research and development work, Dr. Gentile hopes to develop dynamic instructional and assessment tools that support student learning and teachers’ professional development.

James C. Kaufman is the director of the Learning Research Institute and an assistant professor of psychology at California State University at San Bernardino. He is the author or coeditor of six books, including Creativity Across Domains: Faces of the Muse (with Baer; Erlbaum, 2004) and The International Handbook of Creativity (with Sternberg; Cambridge, in press). In 2003, he won the Daniel E. Berlyne Award from Division 10 of the American Psychological Association for outstanding research by a junior scholar. He received his Ph.D. from Yale University and worked at the Educational Testing Service as an associate research scientist for 2 years. He has published more than 65 articles, chapters, and research papers and is the consulting editor of the Korean Journal of Thinking and Problem Solving.

Ann Lupkowski-Shoplik is founder and director of the Carnegie Mellon Institute for Talented Elementary Students (C-MITES). She conducts the annual Elementary Student Talent Search throughout Pennsylvania and oversees the C-MITES summer programs and weekend workshops for academically talented students in ninth grade and younger. Her research interests include identifying mathematically talented students younger than age 12 and studying their characteristics and academic needs. Together with Susan Assouline, she wrote Jane and Johnny Love Math (Trillium Press, 1992) and Developing Mathematical Talent (Prufrock Press,
2003). She is also a coauthor of the Iowa Acceleration Scale with Susan Assouline, Nicholas Colangelo, Jonathan Lipscomb, and Leslie Forstadt.

Dona J. Matthews holds an M.Ed. in counseling psychology and a Ph.D. in special education (gifted) from the Ontario Institute for Studies in Education at the University of Toronto. Since 1985, she has worked in many capacities to support exceptional development in diverse populations, including engaging in academic research, writing, and publishing; consulting with parent, educational, and government organizations; teaching at the graduate level; initiating extracurricular programs for academically gifted children; and conducting a private practice doing psychoeducational assessment and counseling. She is the director of the Hunter College Center for Gifted Studies and Education at City University of New York and the coauthor with Joanne Foster of Being Smart: Understanding and Encouraging Giftedness in Children (in press; for more information, see http://www.beingsmart.ca).

Marcia Strong Scott is a professor of psychology at the University of Miami. Her interests have been in the study of cognitive processes in young children with the aim of using that information to develop screening instruments to identify outliers at both ends of the cognitive continuum: those with mild cognitive problems and those with superior cognitive capabilities. The latter effort is particularly aimed at identifying young cognitively gifted minority children. E-mail: mscott@miami.edu

Mary Ann Swiatek is the research specialist for the Carnegie Mellon Institute for Talented Elementary Students (C-MITES) and a licensed psychologist. She also is an adjunct faculty member in the Psychology Department at Lafayette College. Her research interests focus on the educational and psychosocial needs of gifted youth.

Echo H. Wu completed a systematic research project from 2002–2004 as a M.Phil. student at the University of Hong Kong. She is currently a Ph.D. student in the Curry School of Education at the University of Virginia and a research assistant at the National Research Center on the Gifted and Talented (NRCGT). E-mail: ew5b@virginia.edu