CHAPTER 2

Research Evidence for Domain Specificity

Summary: This chapter explains the kinds of research that can be done to determine the degree to which creativity is domain specific or domain general. It presents the research record, confronts some red herring arguments, and explains what some seemingly conflicting research results really mean.

Introduction

As noted in Chapter 1, we commonly think of some things as domain general, such as intelligence, and other things, such as expertise, we normally think of as domain specific. Even the ways we most often use the terms intelligent and expert in speech reflect these two implicit theoretical positions. We say that Jane is intelligent, and we say that Jack is an expert in early American folk art (or whatever Jack’s expertise happens to be). We don’t say that that Jane is intelligent in something – she is simply intelligent – but when referring to someone as an expert or skillful or knowledgeable, it is common to specify some limited range of skill or knowledge (e.g., an expert chess player or statistician or mechanic, a skillful cook or plumber or writer, or someone knowledgeable about sports or politics or Russian history). People rarely expect others (or themselves) to have expertise or to be skillful or knowledgeable across the board. Such wide-ranging talent or knowledge is sufficiently unexpected that such a person, knowledgeable or skilled in many diverse arenas, is considered unusual and referred to as a Renaissance person. In marked contrast, with intelligence we routinely assume domain generality by simply describing people as intelligent.
(or unintelligent), with the unspoken assumption that this description applies across the board.¹

Common usage would suggest that creativity is, like intelligence, domain general. People often refer to others (or themselves) as “creative” without specifying particular areas or limitations to their creativity. The implicit assumption is that a creative person has some skills, aptitudes, traits, propensities, and/or motivations that lend themselves to creative performance in whatever activities that person undertakes. Creativity is thought of more in the way that intelligence is generally conceptualized, as a general ability that will affect performance in significant ways in almost any endeavor.

The claim that creativity is much more domain specific than commonly assumed, whether by people in general or psychologists in particular, is a relatively new position, but one that has a growing number of adherents. The psychologist who is commonly credited with putting creativity on (or back onto) psychology’s agenda, J.P. Guilford (1950, 1956, 1967; Guilford & Hoepfner, 1971), also argued for a less holistic approach to creativity, but his concern was not primarily with content domains or fields of inquiry. His model of the intellect, of which creativity was a key component, was composed of many discrete (and, he believed, measurable) interacting abilities, but his model was nonetheless virtually silent on the question of domains. It included different skills that led to creative performance, but these skills could, one might assume, be used in many domains where creative performance was possible; it was based on an implicit “assumption of content generality” (Plucker, 1998, p. 178). Any domains in this model were only broadly defined ones such as Guilford’s five kinds of contents (visual, auditory, symbolic, semantic, and behavioral).

The most widely used measures of creativity, the Torrance Tests of Creative Thinking (TTCT), are based on Guilford’s model (Kim, 2006). These tests measure only one component of Guilford’s model, divergent thinking, but they nevertheless claim to predict creative performance generally (Plucker, 1998). The Torrance tests have two forms, the TTCT-Verbal and the TTCT-Figural. The TTCT-Verbal consists of five activities: ask-and-guess,

¹There has been some change in this regard in recent years. The impact of such theories as Gardner’s multiple intelligences (Gardner, 1983) and the even newer focus on emotional intelligence (Mayer, Salovey, & Caruso, 2004) suggests a greater openness to the idea that intelligence may be more domain specific than once thought (or perhaps it simply shows good marketing in using the term intelligence to refer to things that might get less notice if simply termed skills or abilities; although for the purposes of this book, it doesn’t really matter which is the case). Despite strong evidence for a rather significant degree of domain generality in intelligence, the point being made here is not any claim about how domain general or domain specific whatever is being referred to by the term intelligence may be. The important claim here is merely that in common speech people do, unconsciously perhaps, treat intelligence as domain general and expertise as domain specific.
product improvement, unusual uses, unusual questions, and just suppose. The TTCT-Figural consists of three activities: picture construction, picture completion, and repeated figures of lines or circles. This division might suggest a distinction between these two types of creativity, but the two domain-based forms are viewed as different ways to measure the same underlying (and general) construct (Scholastic Testing Service, 2013). Domains don’t matter in most commonly used methods of creativity assessment; the special something that leads to creativity, as assessed by divergent thinking tests (or even the Remote Associates Test; Mednick, 1962; Mednick & Mednick, 1967), is assumed to be the same in all domains.

This chapter reviews the evidence for domain generality and domain specificity. It shows how those favoring one view or the other tend to look for different kinds of evidence to support their views – differences that may suggest different understandings of what it means to be creative – and argues how best to weigh that evidence.2 No reader is likely to be surprised that my conclusions favor domain specificity, but some might be surprised by the arguments and evidence that lead to those conclusions. I also discuss some misleading kinds of evidence that often confuse people when thinking about domain generality and specificity, such as the existence of creative polymaths (which, as we will see, are predicted by both domain specificity

2I believe my presentation of evidence, coming from both sides of this debate, is as even-handed as possible, but as Kuhn (1979) warned us, there is often no neutral ground from which competing theories can be judged or even described. Successive theories are “incommensurable . . . in the sense that the referents of some of the terms which occur in both are a function of the theory within which those terms appear. There is no neutral language into which both the theories and the relevant data may be translated for purposes of comparison” (Kuhn, 1979, p. 409). If this is true, then it is impossible to think in terms of two competing theories at the same time, but only, at best, to switch back and forth between them. Just as one can, in any single moment, only see either a vase or a profile in the vase-profile gestalt found in almost every introductory psychology textbook, theories of domain specificity and generality may be incommensurable. The differences in this case are not so great as between, say, the Copernican and Ptolemaic world views, but in fact such deep revolutions as the Copernican are quite rare; McMullin (1998) showed that most revolutions are “shallow” (p. 122), requiring modification of only small parts of the “disciplinary matrix” that Kuhn (1970) suggested in his postscript to The Structure of Scientific Revolutions are needed to hold together a field of study. But the differences are nonetheless quite real, and within this limited region of creativity’s disciplinary matrix, the distinctions between viewpoints, meanings, and assumptions of domain-general and domain-specific theories can be quite pronounced, and defenders of conflicting theories can easily fail to understand each other’s arguments as a result. “The premises and values shared by the two parties to a debate over paradigms are not sufficiently extensive for that” (Kuhn, 1970, p. 94). I don’t believe what counts as evidence under the two theories (domain specificity and domain generality) need be so very different, however, and I trust that whatever readers’ initial assumptions or beliefs about the domain generality/specificity issue may be, the evidence presented will be both relevant and convincing.
and domain generality, although with different expectations regarding the regularity with which such polymathic creativity should appear).

I have gotten this far without defining creativity. This has been done by design, but the question must be faced. The most common definitions of creativity among theorists and researchers have two parts: (1) creativity involves some degree of originality to the creator (i.e., whatever is being done or thought or produced is not simply a copy of something that one has observed being done or thought or produced); and (2) to some extent at least the idea or product or process must work – it must somehow fit the constraints of the situation as defined by the creator (Plucker, Beghetto, & Dow, 2004). The product or process need not be original in the sense that a patent office might require (i.e., it’s quite okay if the idea is a well-known one, as long as it is new to the creator), and it needn’t solve a problem completely (i.e., it simply needs to move things in the direction of a solution or a positive outcome). I have used definitions of this type myself (e.g., “Creativity refers to anything someone does in a way that is original to the creator and that is appropriate to the purpose or goal of the creator”; Baer & Kaufman, 2012, p. 3).

Definitions of this kind are often helpful and generally harmless, but they assume something generic about creativity that may not actually be especially important when it comes to actual creative thinking or performance in any domain. It seems unlikely that such a definition would be a useful guide in judging, for example, whether a film or joke or a painting or a poem is creative. Experts in a domain judging what is creative in that domain rely on a rich and generally widely shared (among other experts) sense of what constitutes creativity in that domain, and an it-must-be-original-and-workable definition of creativity would be bare bones in the extreme. As a general guide to what might be true of most instances of creativity in most domains, it is probably a reasonable definition, but because it is so generic it leaves out almost everything that is actually interesting or important about actual creativity.

So what is creativity? I can’t offer a better generic definition than the increasingly common it-must-be-original-and-workable one I have just given. But I want to caution that although that may be the consensus definition of creativity in the field, it really isn’t the important definition of creativity. The important definitions of creativity in creativity theory and research – please note the plural definitions – are the ones that operationalize creativity and tell how creativity is measured in a given study, and these vary quite widely. This subject is covered more extensively in Chapter 5, but it is also an issue in considering research related to domain generality/specificity. And as we will see, the it-must-be-original-and-workable construction is rarely central to such operational definitions, just as it is does not usually guide actual determinations of creativity or merit within most domains (Plucker et al., 2004).
Philosophers of science disagree about what constitutes a scientific theory or what kinds of evidence should cause us to favor one theory over another (Curd & Cover, 1998a). The idea that scientific theories must be falsifiable (Popper, 1959, 1963) has probably been the most widely accepted theory among working scientists (Curd & Cover, 1998b). Popper argued that a scientific theory must make clear and testable predictions and that any theory that does not make such predictions is unfalsifiable and therefore not a scientific theory. This standard is appropriate for adjudicating the dispute about the domain specificity of creativity. Fortunately the two sides of this debate do make different key predictions, as summed up by Ivcevic (2007):

Domain generality would be supported by high intercorrelations among different creative behaviors and a common set of psychological descriptors for those behaviors, while domain specificity would be supported by relatively low correlations among different behaviors, and a diverging set of psychological descriptors of those behaviors. (Ivcevic, 2007, p. 272)

Like many questions in psychology (e.g., nature/nurture), the truth in this case may not be all-or-nothing, winner-take-all: creativity might be largely domain specific but with minor domain-general aspects (or vice versa), or the two might be more in balance. But it nonetheless matters to what degree creativity actually is domain specific, because if it is to a significant degree domain specific, it changes how we conceptualize and operationalize creativity, and in doing so it will not only guide us in designing experiments, it will also tell us what past research might (or might not) still be valid. If we assume creativity is largely domain general (which was the standard assumption for many years) but it is in fact largely domain specific, then our measurement techniques and our understanding of creativity may be so far off the mark as to be worth little. Determination of how domain general or domain specific creativity is must, in turn, constrain and guide all creativity assessment.

The two theories of creativity (domain specificity and domain generality) make remarkably different predictions for the most part, and most of the research reported here relates to those predictions and the studies that test those predictions. A few questions about domain generality/specificity do not yield such distinctive predictions, however, especially those relating to the difference between Big-C (domain-altering, genius-level) creativity versus little-c (everyday, garden-variety) creativity. Does the expectation of “high intercorrelations among different creative behaviors” predicted by domain generality mean that Big-C creators would be expected to show high levels of creativity in all (or at least several) fields, or does this expectation only apply to little-c creativity? Or perhaps domain generality would mean that Big-C creators should show Big-C creativity in just one or a few fields along with much higher-than-expected
little-c creativity in other fields – assuming that Big-C and little-c creativity operate by the same processes but at different levels, which is itself a somewhat hard-to-answer question until one has dealt with the domain generality/specificity issue. Those questions will be answered, to the extent that they can be answered, in the section about polymaths later in the chapter.

What Domain Generality and Domain Specificity Predict

To the extent that the skills underlying creative performance are domain-general skills, then those skills should influence creativity on virtually any task one undertakes. Ditto for other domain-general attributes (such as personality traits, thinking styles, or types of motivation) believed to influence creative performance across domains. Domain generality means whatever the factor might be, it has an impact on creativity in all (or at least most) domains. This assumption is made by any domain-general skill-based test of creativity, whatever the form or intended use of the test, or any domain-general assessment of personality, motivation, or thinking styles used to predict creative performance. Those who score high on a domain-general test of creativity should, other things being equal, be more creative than those who score lower on the same test.

Other things are not always (or ever) equal, of course, so other things will also be important contributors to creativity. One may need (or at least would benefit from having) specialized domain-specific skills and knowledge that might be combined productively with the skills hypothesized to underlie domain-general creativity to produce more creative outcomes. Interest in working in a specific domain, the availability of domain-specific tools, and other domain-specific factors may also influence the level of creative performance in a given domain. But if creativity is domain-general, then on average, people who have more of whatever those creativity-relevant domain-general skills or attributes happen to be will be more creative than those who have less of those skills or attributes.

Defining and measuring those domain-general skills or attributes have been more difficult than one might hope, nor is there consensus on what those skills and attributes might be. But a much simpler, clearer, and more universal prediction of domain generality flows directly from the theory of domain generality and can be tested more directly, without the need to agree on the nature of the specific contributors to creativity, be they skills, personality attributes, motivations, or thinking styles. It is a prediction that does not require any specification of what the skills or attributes underlying domain generality might be or any agreement among domain generality theorists beyond the most basic core of the theory. That prediction is this: To the extent that creativity is domain general, people who are more creative than most other people in one domain should be more creative
(other things being equal) than most other people in other domains as well. This statement parallels the primary argument for the existence of $g$ in the intelligence literature. The prediction thus is a straightforward one: Domain generality predicts positive correlations between creativity measures or tests given to the same person that assess creativity in different domains. The stronger the correlations across domains, the more evidence for domain generality. Domain specificity predicts the opposite: little or no correlation between scores on creativity tests or other measures of creativity drawn from different domains.

Domain generality of creativity thus predicts, at a modest to high level, positive correlations among the creativity ratings of artifacts produced by subjects in different domains. Domain specificity predicts the opposite: low or nonexistent levels of correlation among creative products produced by subjects in different domains. Advocates of domain specificity in the area of creativity who accept a domain-general intelligence factor and who believe that intelligence is one factor influencing creative performance would predict a low level of correlation among the creative products produced by subjects in different domains. This correlation is caused by domain-general intelligence that affects performance across domains (and to the extent that IQ tests measure $g$, the degree of influence of $g$ on creative performance would be measurable by those tests, and that impact could then be statistically removed by parceling out variance attributable to $g$). Most who argue for domain specificity take this position: a prediction of low correlations on different domain-based measures of creativity, with those low correlations explainable by differences in intelligence (Baer, 1993, 2010, 2013).

I noted above that skills (whether domain-general or domain-specific) aren’t the only theorized contributors to creativity. Many other factors might contribute to domain generality, such as motivation and conscientiousness. Motivation might, of course, be domain specific, and to some degree it certainly is (e.g., someone who finds history fascinating might not have similar motivation to study chemistry or philosophy). But motivation might also be a fairly general, domain-transcending attribute, one that would influence performance in any domain. If so, this tendency should show up in across-domain correlations of creative performance. Similarly, traits such as conscientiousness might be either domain-transcending (so that a person who is conscientious when tending a garden or editing a manuscript would also tend to be conscientious when doing cooking, child care, or accounting); or one might have varying levels of conscientiousness depending on the domain with little commonality across domains, making conscientiousness much more domain specific.

To the extent that a person acknowledges any domain-general factors influencing creative performance across domains, that person would expect to see higher correlations among the creativity ratings of products produced by subjects in different domains, whatever the factors.
Advocates of domain specificity who do not accept any general factors (intelligence, motivation, or conscientiousness) that might affect creative performance across domains would predict zero or random correlations among those ratings.

One further complication could arise. It might be that some traits that influence creative behavior are domain-general but nonetheless have a domain-specific impact on creativity. Consider, for example, the trait of conscientiousness. There is rather convincing evidence that in the case of conscientiousness, as measured by tests of the Big Five Personality traits, there is a considerable degree of domain generality. If one assumes conscientiousness is a domain-general trait, then a person who is conscientious in artistic or literary pursuits would also be expected (other things being equal) to be conscientious in doing mathematical or scientific activities. But what if conscientiousness had a positive impact on creativity in some domains but a negative impact on creativity in others domains (and perhaps no impact on creativity in still other domains)? In such a case, even a domain-general trait would have a domain-specific impact on creativity. The example of conscientiousness was offered hypothetically, but evidence from personality testing indicates it is a domain-general trait and evidence from creativity research suggests that conscientiousness has a significant positive impact on creativity in some domains (such as some scientific fields) and a significant negative impact in others (such as some artistic fields; Feist, 1998, 1999). This example reveals how even a seemingly domain-general trait might lead to domain specificity in creativity.

Summing up:

1. There are many candidate skills, traits, and dispositions that might influence how creative someone is.
2. Those skills, traits, and dispositions might themselves be either domain general or domain specific.
3. Even among domain-general skills, traits, and dispositions, it is possible that some might have a domain-specific impact on creativity (e.g., conscientiousness, which may have a positive impact on creativity in some domains and a negative impact on creativity in other domains).
4. Any skills, traits, or dispositions that support any kind of domain-general theory of creativity should lead to significant positive correlations between creative performances across domains. The higher the correlations, the stronger the domain-general effect.
5. Domain specificity predicts low (or zero) correlations between assessments of creative performances across domains.
6. Intelligence is considered a domain-general factor that influences creativity across domains even by many domain specificity theorists,
who therefore predict low correlations among creative performances across domains that can be accounted for by differences in intelligence (such as is measured by tests of \( g \)).

The picture presented here is a complex one (which is hardly surprising; after all, it is creativity we are trying to understand!). Fortunately, the most direct test of the domain generality/specificity question avoids these complications by looking (as noted above) at the most basic prediction that divides the two theories: Domain generality predicts positive correlations between creativity measures or tests given to the same person that assess creativity in different domains. The stronger the correlations across domains, the more evidence for domain generality. Domain specificity predicts the opposite: little or no correlation between scores on creativity tests or other measures of creativity drawn from different domains.

**What Cross-Domain Creativity Assessments Show**

The preceding discussion leads to a rather simple test: Are people who are more creative in domain X also more creative in domains Y and Z; that is, are “high intercorrelations among different creative behaviors” (Ivcevic, 2007, p. 272) present, as domain generality predicts? Assessment of creativity is difficult, unfortunately, and most of the methods commonly used have approached creativity rather indirectly, via skills or behaviors or traits that are theoretically linked to creativity. It is rather like using height as a measure of basketball skill. Height is probably positively correlated with basketball skill, but using it as a proxy for basketball skill would provide a meager measure of such skill. Such indirect methods are also problematic because they require empirical links to actual creative performance, which in the end must be what validates any creativity measure. To continue the basketball-height analogy, one would have to show that height is indeed correlated with basketball skill, but if one can do that, then why would one jettison the criterion measure of actual basketball skill and replace it with the proxy of height?

The same problem plagues indirect measures of creativity based on theoretical connections to actual creativity. Conceptual connections to creativity can certainly be useful guides and creativity theory can help in the search for potentially useful measures, but all such measures must in the end be grounded in actual creativity. Creativity test scores must be correlated with actual creative performance; creative performance cannot be free-floating. Because creative performance occurs in domains, it must be observed and measured in domains. It may occur in more than one domain, of course (e.g., musical theater calls for creativity in many domains), but creativity cannot occur in no domain at
all. Whether a tangible product or an ephemeral idea, creativity cannot occur in a vacuum without ties to at least one domain, field, discipline, or area of interest. Measures of creativity are therefore also linked to domains. A domain-specific measure of creativity must be correlated with creative performance of some kind in the given domain, and a domain-general test of creativity must be correlated with creativity across many domains, not just a few.

The most common measures of creativity are divergent-thinking tests, but these tests have not been helpful in the generality-specificity debate. For the most part, such tests assume domain generality, and therefore all that most standard divergent-thinking tests report are domain-general scores. One could look for correlations between such test scores and creativity across many domains, and attempts to do this as ways of validating divergent-thinking tests have been tried (with little success; reports of what we know about research results in this area follow). But as a way of adjudicating the domain generality/creativity dispute, looking for these correlations just adds an unnecessary intermediary that can only lower estimates of domain generality. Unless the test is perfect, the error variance of the divergent-thinking test scores just becomes an added source of error. If one has measures of creative performance in many domains (which is what one would need to validate a domain-general divergent-thinking test), then there is no reason to look at the test scores at all if one’s interest is domain generality/specificity. The correlations that matter are the ones among the various measures of creative performance in different domains. Domain-general divergent-thinking test scores (or scores on any domain-general test of creativity), even if the test is valid, could add nothing except additional error variance to such calculations.

Domain-specific divergent-thinking tests are also quite possible; in fact, Torrance (1966, 1974) himself made a step in this direction with his verbal and figural forms of the TTCT, although he believed they were both testing a single domain-general skill that happened to be deployed in different domains, despite the fact that he found almost no correlation between scores on his two divergent-thinking tests (Cramond, Matthews-Morgan, Bandalos, & Zuo, 2005). To the extent that such domain-specific tests are valid, they could be used to show commonality (or lack thereof) in divergent-thinking scores in different domains. Such use would first require not only the creation of tests of domain-specific divergent-thinking skills, but also separate validation of each domain-specific divergent-thinking test. This type of test and validation has not been done, and Torrance’s failure to find positive correlations between scores on his two different types of divergent-thinking tests (which although intended to be domain-general tests nonetheless used content from two different domains because, once
again, creative performance must occur in some domain) does not bode well for efforts by anyone hoping to establish domain generality based on such an approach. What little we know about correlation between divergent-thinking tests in different domains argues for the domain specificity of divergent thinking (Clapham, 2004; Cramond et al., 2005).

There is one additional problem with using divergent-thinking tests to look for domain generality. Such tests are at best measures of a skill or set of skills – divergent-thinking skills – that although theoretically linked to creativity are nonetheless, at most, just one aspect of creativity, and therefore not actually a full measure of creativity itself (just as height is not a full measure of basketball skill, as in the analogy above). Even if one assumes that creativity is domain general and that divergent thinking is a component of creativity, calling divergent-thinking tests creativity tests would be rather like calling tests of one’s ability to recall strings of random numbers intelligence tests. At most they would be but one part of a larger general factor. Personality and trait theories of creativity also most often assume domain generality, and once again these are not measures of creative performance but rather things that are either theoretically or empirically linked to creativity (Kaufman, Plucker, & Baer, 2008b). None of these standard methods of creativity assessment is appropriately free of theoretical bias and none measures actual creativity, only some limited range of its surrogates that are believed to be correlated with creativity.

There is, however, one method of creativity assessment that quite well suited to test the domain specificity question. The Consensual Assessment Technique (CAT), a method of assessing creative performance (most often of some creative product), was originally developed by Teresa Amabile (1982, 1983, 1996) and further developed by others (e.g., Baer, Kaufman, & Gentile, 2004; Hennessey, Kim, Guomin, & Weiwei, 2008; Kaufman, Baer, Cole, & Sexton, 2008a). It measures creativity directly, not indirectly, by asking experts in the field in question to rate the creativity of a group of artifacts. It works so well that it has been called the “gold standard” of creativity assessment based on the following three arguments (Carson, 2006):

1. Because it is based on evaluations of actual creative products, performances, or artifacts, it is a measure of the actual creativity of those products, not just of things believed to be related in some way to creativity.
2. It isn’t linked to any particular theory of creativity or dependent for its validity on any such theory.
3. It employs the same method for assessing creativity that is used in most domains in the real world. No rubric or other intermediary lies between the experts in the field and their ratings of creativity.
of work in that field. In any domain, the best estimate of the creativity of work in that domain is the combined judgment of experts in that domain. Such expert judgments may, of course, change over time – later experts may value a given work more or less highly than their predecessors – but at any point in time, no assessment of creativity in a domain is more valid than that of its combined experts.

The CAT simply and directly asks experts to rate the creativity of products in a domain in comparison to one another, in the same way that, say, the Academy of Motion Picture Arts and Sciences asks experts in the field to rate movies, actors, and directors, or Nobel Prize committees in different fields rate the work of practitioners in their fields. The CAT is certainly not perfect (neither, one could argue, are the judgments of Nobel Prize committees), but it is perhaps the best available method to assess real-world creativity. As a recent handbook of creativity assessment concluded:

The CAT is based on this idea that the best measure of the creativity of a work of art, a theory, or any other artifact is the combined assessment of experts in that field. Whether one is selecting a poem for a prestigious award or judging the creativity of a fifth grader’s collage, one doesn’t score it by following some checklist or applying a general creativity-assessment rubric. The best judgments of the creativity of such artifacts that can be produced – imperfect though these may be – are the combined opinions of experts in the field. That’s what most prize committees do (which is why only the opinions of a few experts matter when choosing, say, the winner of the Fields Medal in mathematics – the opinions of the rest of us just don’t count). The CAT uses essentially the same procedure to judge the creativity of more everyday creations. (Kaufman et al., 2008b, pp. 54–55)

When using the CAT, the experts rate the creativity of a set of artifacts by comparing them to one another. Each expert judge works independently of all other judges and there is no opportunity for them to influence or even know anything about other judges’ opinions, bases for making ratings, or the actual ratings of creativity other judges have given to any of the things being judged. They have no opportunity to discuss the artifacts to be judged, nor do they even know, in most cases, who the other judges might be. They are given no instructions other than to use their own expert sense of what is creative in a domain.

The Hungarian mathematician Paul Erdos explained that one cannot explain the beauty of numbers (or, I would argue, mathematical creativity), one can only appreciate it (assuming one has the requisite expertise). There is no rubric, no external standard, by which either beauty or creativity can be explained or verified outside of the domain itself. “It’s like asking why Beethoven’s Ninth Symphony is beautiful. If you don’t see why, someone can’t tell you. I know numbers are beautiful. If they aren’t beautiful, nothing is” (Erdos, quoted in Devlin, 2000, p. 140).
Despite working entirely independently, the interrater reliability among the ratings the judges give is quite good, generally in the 0.80–0.90 range (Amabile 1982, 1983, 1996; Baer, 1993; Baer et al., 2004; Kaufman et al., 2008b). These consistent results attest to the reliability of the CAT. It is the use of expert judges, and complete reliance on their expert judgments, that assures the validity of these creativity ratings.

A number of studies looking for domain generality have been conducted using CAT and CAT-like assessments of the creativity of subjects in a variety of domains. The results have been consistent: low intercorrelations among the creativity ratings of different artifacts produced by the same subjects. In one typical study, Baer (1993) asked 50 eighth-grade students to create poems, stories, mathematical word problems, and interesting equations in which students were asked to create a mathematical equality that they considered especially interesting (see Baer, 1993, pp. 49–52, for more complete details on the tasks). Each student created one of each type of artifact. There were two groups of 25 participants, with the order of the tasks reversed for the two groups.

Students’ responses to these four creativity tests were typed and photocopied by the experimenter and then rank ordered for creativity by five qualified experts, who were paid for their work. The expertise of the judges depended on the test. For example, the poems were rated by poets and English teachers, and the equations were judged by mathematics teachers and mathematics professors. (Baer, 1993, p. 50)

To make the task manageable, the rank-ordering was done of just 25 papers at a time, so each judge on each task needed to deal with 25 papers (so there were two papers at each level: two with a rating of most creative, two with a rating of second-most creative, and so on down to two papers that each were rated 25th). The interrater reliabilities (coefficient alpha; Nunnally, 1978) of these four sets of 50 papers were 0.78 (word problem), 0.86 (poems), 0.89 (stories), and 0.92 (equations).

Of the six cross-domain correlations, half were positive and half were negative, with a mean correlation of 0.06 and with just one of the six reaching statistical significance, as shown in Table 2.1.

**TABLE 2.1** Correlations Among Creativity Ratings

<table>
<thead>
<tr>
<th>Task</th>
<th>Poetry</th>
<th>Story</th>
<th>Word problem</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poetry</td>
<td>–</td>
<td>0.23</td>
<td>0.31*</td>
<td>−0.14</td>
</tr>
<tr>
<td>Story</td>
<td>−</td>
<td>−</td>
<td>0.20</td>
<td>−0.03</td>
</tr>
<tr>
<td>Word problem</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−0.20</td>
</tr>
</tbody>
</table>

*N = 50.*

*p < 0.05, two tailed.*
In this study it was hypothesized that general intellectual ability might add to cross-domain correlations, and so variance attributable to math and verbal standardized test scores was removed. One again the result was three positive and three negative correlations, this time with a mean correlation of –0.05. The only statistically significant correlation found after variance attributable to measures of general intellectual ability had been removed was a negative one, as shown in Table 2.2. (It should be noted that students in this sample all had above-average scores on these two standardized tests. This similarity among the participants on these tests might have minimized the effect of general intellectual ability on creative performance. This is not important for the assessment of domain specificity or domain generality of creativity unless one is arguing that general intelligence is the domain-general factor, but that is expressly not the argument made for domain generality. The argument for domain generality is that a domain-general factor influencing creativity across domains is distinct from general intelligence.)

Baer (1991, 1993, 1994b) reported similar results with adults, fifth-grade students, fourth-grade students, and second-grade students, which together made a “strong case for an absence of any significant effects of general creative-thinking skills on the performance of a wide range of subjects on a variety of creativity-relevant tasks” (1993, p. 67).

Other researchers have produced comparable results using different tasks and populations. For example, see Han (2003) and Han and Marvin (2002) conducted a study using 109 second-grade students as subjects:

Three performance-based assessments were utilized in this study. The three assessments included: a story-telling task (language), a collage-making task (art), and a math word-problem task (math). All tasks were selected from various sources using Amabile’s (1983, 1996) guidelines for selecting appropriate tasks for a consensual assessment. . . . The story-telling and the collage-making tasks for the present study were developed by Amabile (1983) and have been validated in more than 30 independent studies by Amabile and her colleagues (1983, 1996). (Han, 2003, pp. 122–123)
In this study of story-telling, collage-making, and mathematical word problem-creating creativity, Han (2003) reported just one statistically significant ($p = 0.04$) correlation across domains (story-telling and math problem; $r = 0.283$, accounting for 8% of total variance). No measures of general intellectual ability were used in this study, so the influence of general intelligence, which may have accounted for some of that variance, could not be assessed. Han also gave participants two verbal divergent-thinking subtests (Alternate Uses and Similarities) and one nonverbal subtest (Pattern Meanings) of the Wallach-Kogan Creativity Test and a real-world divergent-thinking test (adapted from Okuda, Runco, & Berger, 1991) and reported that no combination of these various tests and subtests “could explain or predict significantly any combination of the three performance-based assessments.” Han noted a statistically significant correlation between two subtest scores and the storytelling task, a finding that was “consistent with some previous research in that divergent thinking measures were often weakly but significantly related to creative activities in the language domain . . . implying a possible verbal bias in divergent thinking tests” (p. 129). Han also reported results of a case study that aligned closely with the quantitative results reported here.

Working with an older population, Ruscio, Whitney, and Amabile (1998) asked undergraduate subjects to complete three tasks (structure-building, collage-making, and poetry-writing) and found little evidence of general creativity (correlations of 0.18, 0.09, and –0.02 across domains). Of these, only the correlation between structure-building and collage-making reached the 0.05 level of statistical significance ($r = 0.18$, accounting for a little more than 3% of the total variance). These two tasks (structure-building and collage-making) may not actually be from two distinct domains but rather from the same general thematic area. The instructions for one were to “build an aesthetically appealing structure that’s at least fifteen inches tall” (p. 248) and for the other to “make a collage out of the materials you see in front of you” (p. 249). Both asked subjects to create a work of art, one in three and the other in two dimensions. It is difficult to interpret these results as evidence of domain generality because the only commonality found among the creative performances of subjects on different tasks was between tasks that could be seen as coming from the same general thematic area. No positive correlations of measures of creative performance across domains were found. Similarly, in a study using expert raters to assess the creativity of elementary school children’s art, Runco (1989) found low correlations (median $r = 0.18$) among the different kinds of works of art produced by his subjects. Even within the same broadly defined domain of art, there was only a modest degree of generality across different tasks, which is similar to the results given by Ruscio et al. (1998) results, with no evidence of truly domain-general creativity.

There is one study (Conti, Coon, & Amabile, 1996) that asked participants to create a number of different artifacts, used the CAT to assess the creativity of those artifacts, and claimed their results supported Amabile’s
componential model of creativity, which posits both domain-specific and creativity-general skills that influence creative performance. (These two kinds of factors are in addition to a third factor, task motivation, which has been Amabile’s primary interest and focus of study; see Amabile, 1983, 1996.) The study by Conti et al. was actually a reanalysis of data that had been previously collected as part of three different studies. This combination was possible because some subjects in those previous studies had participated in two or all three of the studies and therefore correlations among their creative performances on a number of different tasks in two domains could be computed.

The subjects of Conti et al. (1996) completed a total of four story-writing tasks (using different prompts) and three different art activities. The intercorrelations they reported among the story-writing creativity ratings were indeed both high and statistically significant, suggesting that these measures were largely measures of the same domain-based ability. For example, the intercorrelations among the three stories written as part of one study ranged from 0.43 to 0.87, confirming the prediction that “creativity measures taken within the same context and domain should be strongly positively related” (p. 387). Correlations with creativity ratings of these stories and stories written at a different time and under different experimental constraints were, as expected, somewhat lower; “as predicted, creativity measures within the same domain are substantially intercorrelated, although not as strongly as those taken within the same experimental context” (p. 387). It should be noted in passing that this finding (of significant, but somewhat weakened, correlations in creativity ratings on similar activities over extended periods of time between testing) is in accord with previous studies that have shown substantial correlations between ratings of creative activities in the same domain over time. For example, Baer (1994c) found fairly robust long-term stability using essentially the same short story-writing task (albeit with different prompts) with a 1-year interval between testing. The story-writing creativity of 9-year-old participants correlated 0.58 with the story-writing creativity of the same participants 1 year later, which is not far off the 0.60–0.80 stability coefficients found for IQ test scores at this age (Kogan, 1983). Baer also found significant, but somewhat smaller, correlations in poetry-writing creativity over the same 1-year interval.

Conti et al., 1996 reported that the correlations among the ratings of the art-related tasks were also positive, but not as strong, because unlike the writing tasks, which were all similar (all required subjects to write a short story based on a prompt), the art tasks varied considerably from one to the next. “Here the tasks were substantially different, stretching the definition of ‘domain’ somewhat. Nonetheless, drawing and collage creativity are highly correlated, and painting and collage creativity are moderately correlated” (p. 387).
These correlations are all within-domain correlations, of course, and both domain generality and domain specificity predict one would find such correlations. They show that creativity on different tasks in the same domain is highly correlated, and the more closely related the tasks are, the higher the correlations. These correlations tell us nothing, however, about the domain specificity/generality question. It is not the within-domain correlations but the cross-domain correlations that speak to the generality/specificity question, and here the results were different. Of the 13 correlations of this kind, eight were positive, four were negative, and one was zero. None of these 13 correlations was statistically significant, which means they provided no substantive evidence at all for domain generality. The mean value of these 13 correlations was 0.109, which would account for barely more than 1% of the variance. No measures of general intellectual ability were reported, but based on the results reported by Baer (1993) when such measures were used to remove variance attributable to such factors, it is likely that the results would have been even more dismal for domain generality.

One recent study did report evidence of domain generality across artifacts in different domains. Chen, Himsel, Kasof, Greenberger, and Dmitreviva et al. (2006) wrote that in contrast to all previous studies of this type, which had consistently shown no evidence of domain generality, theirs was “the first study to our knowledge that provides reasonable psychometric evidence” (p. 195) for the domain generality of creativity. Their subjects were 159 undergraduates, each of whom produced a number of products in different domains. A principal components analysis of the creativity ratings resulted in three factors that generally corresponded to the domains of artistic, verbal, and mathematical creativity, which were the three kinds of tasks the subjects performed. Using these they created three summary scores of subjects’ verbal, artistic, and mathematical creativity and submitted these scores to a factor analysis and extracted a single factor that accounted for 45 and 52% of the variance in two subject groups.

These results are different, as the authors noted, from all previous research of this kind, but they were marred by a crucial failure to follow the required procedures for the CAT. Rather than use expert judges, which is the basis of the CAT’s validity, they replaced the required expert raters with “trained undergraduate research assistants” (Chen et al., 2006, p. 186). As explained above, the use of expert judges is the (sole) basis for the CAT’s validity claims, and the substitution of novices is not supported either (1) by Amabile’s (1982, 1983, 1996) original work on the CAT, where she wrote that “it would be a mistake to conclude that everyone (or even every psychology graduate student) can be considered an appropriate judge” and “the best guideline is to use judges who have at least some formal training and experience in the target domain” (Amabile, 1996, p. 72); or (2) by more recent work comparing the ratings of novices (college
students like the ones used by Chen et al. (2006) with experts. In the domains of poetry (Kaufman, Baer, Cole, & Sexton, 2008) and short stories (Kaufman, Baer, & Cole, 2009b), experts’ creativity ratings and the creativity ratings of undergraduates were not sufficiently correlated to allow the replacement of expert judges by novices (and these studies used two of the task domains employed by Chen et al., whose subjects created two poems and one story). This finding – that novices like college students could not validly replace experts when using the CAT – was recently confirmed by Kaufman, Baer, Cropley, Reiter-Palmon, and Sinnett (2013a) in a study using both artistic and engineering tasks. (For a summary of findings about the use of nonexpert judges and the CAT, see Kaufman & Baer, 2012.)

Because they can make no claim of validity for their data, Chen et al. (2006), in what they claimed was the first and only study to provide reasonable psychometric evidence for domain generality of creativity, in fact produced no findings that support domain generality.

Other researchers using the CAT have looked for evidence of domain generality in a different way – by training participants in domain-specific divergent-thinking skills and then assessing creative performance in that domain and other domains (Baer, 1994a, 1996). These studies have shown that it is only creativity in the area of the training (where the domain used for the training exercises matches the domain of the creativity testing) that shows an increase in creativity posttesting. Creativity ratings on tasks in other domains or subdomains (ones that are different from those used in the training exercises) are not affected by domain-specific creativity training. This is not the same kind of direct test that the studies employed (it is more closely relevant to the questions addressed later in the chapter on creativity training), because not even the most ardent advocate of domain generality would claim that there are not also domain-specific skills that contribute to creative performance, but they do add to the mountain of evidence from CAT-based studies that have found little or no evidence of domain generality.

**Validity Check Interlude**

The kind of research reported in the previous section was initially challenged by Kogan (1994), who argued that limited sample size and restriction of range may have limited the size of the observed correlations in some of Baer’s (1991, 1993) early studies. For example, in the study of 50 eighth-graders’ creativity in four different tasks, all participants were in the upper quartile academically. A partial replication of that study was therefore conducted, this time with the entire eighth grade (N = 128) of a middle school with an academically diverse population (Baer, 1994b). Just two tasks, poetry writing and story writing, were used; these two had one
of the highest correlations (0.23) reported in the earlier study, which might be expected because they come from the same general thematic area of writing (although from different domains within that field). In the 1994 replication, this correlation actually dropped slightly (to 0.19), suggesting that the design of the earlier study had not prejudiced the results.

Because CAT assessments are essentially single-item tests (i.e., subjects generally produce one story, one collage, one poem, etc., not 10 stories, 10 poems, or 10 collages), one might also question the reliability of the assessments used, which is an issue with any assessment technique. Interrater reliabilities are generally quite high, as noted earlier, but intrasubject reliabilities could be lower, and low reliability would artificially reduce the observable intercorrelations. Like any assessment, CAT assessments are not perfectly reliable, and a correction for attenuation can be used to estimate the extent to which observed correlations are attenuated by measurement error (Cohen & Cohen, 1983; Nunnally, 1978). To the extent that measurements are unreliable, correlations between those measures will be reduced, but an estimate can be made of what the correlations would have been if perfectly reliable measures had been used. Some controversy remains about when or if this correction should be applied (Cohen & Cohen, 1983; Nunnally, 1978), but even if used it makes little difference in the data presented here regarding intercorrelations across domains. The impact of this correction increases with the unreliability of the measures, and as the reliabilities for the most part are quite good (typically in the 0.80–0.90 range), the impact is small. The magnitude of the effect also increases with the size of the correlation, however. This means that the much higher correlations found among creativity ratings of artifacts in the same general thematic area or domain increase more when corrected for attenuation than the low or nonexistent correlations of creativity ratings across domains. The effect on the interpretation of the results is minimal; the changes produce slightly larger positive and slightly larger negative correlations. Little change occurs in the overall pattern, or in the general conclusion that little evidence supports the influence of domain-general creative-thinking skills such as divergent thinking. Baer (1993) reported corrections for attenuation in all seven of the studies he presented. Here are all

4CAT-rated assessments of creativity are basically single-item tests, but the “single item” is not a brief response like a multiple-choice answer. Each single item being judged by experts in CAT assessments is a complete and often complex product, such as a short story or a collage, which provides a much richer assessment even though there is but a single thing to judge. Similarly, comparisons of novels submitted for a contest such as the Booker Prize or of films for a Directors’ Guild Award also involve single-item tests—each novel or film being a single item—but these are also content-rich single-item “test” materials.
the changes based on correction for attenuation for the study of eighth-grade students reported earlier:

<table>
<thead>
<tr>
<th>Tests</th>
<th>From</th>
<th>To</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poetry – story</td>
<td>0.23</td>
<td>0.26</td>
</tr>
<tr>
<td>Poetry – word problem</td>
<td>0.31</td>
<td>0.38</td>
</tr>
<tr>
<td>Poetry – equation</td>
<td>-0.14</td>
<td>-0.16</td>
</tr>
<tr>
<td>Story – word problem</td>
<td>0.20</td>
<td>0.24</td>
</tr>
<tr>
<td>Story – equation</td>
<td>-0.03</td>
<td>-0.03</td>
</tr>
<tr>
<td>Word problem – equation</td>
<td>-0.20</td>
<td>-0.24</td>
</tr>
</tbody>
</table>

How, then, does domain-specific expertise and creative performance affect the assessments used in these studies? In all these assessments, current creative performance is assessed, not some other variable that might predict future performance. As such, the CAT can be thought of as being more like an achievement test than an aptitude test; its goal is simply to measure current levels of creativity. In most cases, subjects with more experience in a domain are likely to evidence more creativity in that domain because domain-specific prior knowledge and experience matter in creativity. In most domains such domain-specific skill and knowledge are part of what one needs to be creative in a given domain, even under domain generality (which typically posits that there are both some domain-specific factors like content knowledge or skills and significant domain-general factors that influence creative performance across all domains). One might therefore worry that domain-specific differences in subjects’ prior knowledge and experience could undermine the use of the CAT to test for domain generality, so it’s important to explain why that is not a problem.

The developer of the CAT, Teresa Amabile (1982, 1983, 1996), used fairly common tasks such as collage-making and story-telling that required little formal training because her main interest was in changes resulting from different motivational constraints, although she understood that training and experience would still influence creative performance even with these familiar tasks. One wouldn’t want to assess subjects’ creativity in a domain that is totally unfamiliar to them (e.g., asking everyone to write a concerto, or to write a haiku in Japanese) because that would result in few or no subjects producing anything. But as long as all subjects had some experience in the domain, she didn’t think the fact that some subjects might have more knowledge and experience would be a problem. Her goal was not to assess some hidden, possibly innate but undeveloped
creative ability. The focus of the CAT is on current levels of creative performance, not what subjects might have done (or might be able to do) with proper training.

Let me illustrate this distinction with a quote from *Pride and Prejudice* (Austen, 2008/1813) in which Lady Catherine de Bourgh made an unintentionally comical appeal of this kind, one that tried to shift the focus from achievement to aptitude when she argued that if she and her daughter had only had musical training, they would have been quite proficient:

There are few people in England, I suppose, who have more true enjoyment of music than myself, or a better natural taste. If I had ever learnt, I should have been a great proficient. And so would Anne, if her health had allowed her to apply. I am confident that she would have performed delightfully. (p. 195)

It is possible (but unlikely) that Lady Catherine and Anne might have great hidden musical talent and that training could have brought forth extraordinary results, but in the meantime, neither is proficient (or creative) musically. Similarly (and like the assessments made by almost all award committees, most of which use a process similar to the CAT), what the CAT assesses is what someone can do now, employing only the skills, knowledge, interests, motivations, and so on that they bring to a particular task at a particular point in time.

One’s level of creativity in a domain may change as one gains more experience, skills, interest, or knowledge in a field, of course. As explained above, research has shown that, in general, CAT ratings in a given domain are fairly stable over time; subjects whose work has received CAT-based ratings in a domain tend to receive similar ratings when tested using a different task in the same domain a year later. For this reason, CAT-based ratings do predict future creative performance on particular kinds of tasks, such as writing stories or poems, rather well (Baer, 1994c), and can therefore be used validly when one needs domain-based predictors of future creative performance (Baer & McKool, 2009, 2014). But these ratings have also been shown to change substantially with domain-specific training, although only in the domain where training has occurred (Baer, 1994c, 1996).\(^5\)

\(^5\)It should also be noted that, unlike divergent-thinking tests, it is impossible to game CAT measures. The scoring rules for divergent-thinking tests can be explained quickly, and knowing how a divergent-thinking test will be scored (e.g., knowing that simple fluency—a large number of responses—will boost scores) makes it possible to increase one’s score fairly easily, and with just a little training divergent-thinking test scores can be inflated based on that training. “After a testing session in which teachers were being trained to score the Torrance, one test taker said, ‘I could have done so much better if they’d just told me ahead of time what they were looking for’” (Mensa, 2015). This is not true with CAT ratings.
Because current levels of creativity in a domain will vary with levels of skill and knowledge in that domain, which are both things that will often reflect training and experience that some subjects might have had greater opportunities to acquire, it is possible that one might find more domain generality when using CAT ratings to test for domain generality if all subjects had similar levels of training and knowledge in all fields. This is why using fairly young children with similar educational backgrounds and using common tasks with which everyone has at least some experience may be the best way to conduct these tests, although then one gets complaints similar to Kogan’s (1994), noted above, about possible restriction of range. But even if this is a valid concern (and it is), it does not change the results reported or their interpretation, because if creativity were a domain-general skill then it should still tend to heighten creativity in all areas (just as a rising tide will lift all boats, large and small), so even when using adult subjects with widely varying expertise in the domains in question, domain generality would still predict significant positive correlations across domains. That is not what research has found, however, with either children or adults. The observed correlations simply don’t support much in the way of domain generality.

Summary of What Assessments of Actual Creative Products Tell Us About Domain Generality

In sum, all available evidence from the many studies that have looked at actual creative products in search of the “high intercorrelations among different creative behaviors” (Ivcevic, 2007, p. 272) that would demonstrate domain generality suggests that creativity is largely domain specific. What these studies have typically found is either low or essentially random correlations. Based on studies that looked at actual creative products, it is difficult to escape the conclusion that creativity is overwhelmingly domain specific. There are other kinds of evidence (reviewed in the next section) that some theorists have argued make a case for domain generality, but what most would agree is the best kind of evidence – actual creative performance on real creativity-relevant tasks – leaves little room for domain generality beyond what might be attributable to general intelligence.

This evidence for the domain specificity of creativity does not mean that domain specificity theorists have reached consensus on what the primary domains might be. Feist (2004) proposed seven “domains of mind”: social-emotional, physics, natural history, language, mathematics, art, and music. These categories are somewhat similar to Gardner’s (1983, 1999) well-known eight intelligences (language, logical-mathematical, interpersonal, intrapersonal, spatial, natural history, bodily-kinesthetic, and musical). Feist cataloged six other, somewhat similar, domain inventories and provided evidence that his seven domains of mind are fairly universal.
Perhaps, but it would be premature to suggest that most domain-specificity theorists think the domains of creativity have been clearly mapped. Much more work remains to be done in this arena.

Establishing domains is a separate issue from showing that creativity is domain specific, however. To map domains requires domain specificity, but domain specificity does not require a mapping of domains. It may be true that rather well-defined domains of the kind Feist and Gardner have proposed do exist, but it is also possible that domain boundaries are much less distinct, and clear-cut divisions may not even exist. It seems likely that even within domains, some tasks call on more similar creativity-relevant skills than others, so even within domains, subdivisions as well as possible subdivisions within subdivisions are likely. The evidence presented indicates that even within a domain, creative performance may vary significantly depending on how similar the tasks are (e.g., Conti et al., 1996, in which much higher intercorrelations were noted among the creativity ratings of the writing activities, which were similar to one another, than among the art activities, which came from different subdomains within the larger domain of art). This result suggests that what may be more important is the similarity of two tasks rather than whether those tasks come from the same domain, which would make labeling the domains (and subdomains) both more difficult and perhaps less crucial. In Chapter 7 of this volume, I offer a hierarchical model that includes both domain-general and domain-specific features, with empirically derived domains that are similar to Feist’s but with other levels of a hierarchy both above and below that of domains. But the accuracy of that or any model of the domains of creativity is not essential to making the key point that creativity is largely domain specific.

Evidence for Domain Generality

The idea that cognitive abilities, particularly creative ones, are content general is currently much maligned: Creativity and other thinking skills applied within certain content areas are widely believed to be independent of creativity and thinking skills applied in other content areas. . . . [One] could reasonably assume that the debate is settled in favor of content specificity. (Plucker, 1998, p. 179)

These are the words with which Plucker (1998) opened his Point-Counterpoint article in the Creativity Research Journal about domain generality/specificity – and he was the one arguing for domain generality! He acknowledged that the tide of evidence seemed to have turned in favor of domain specificity, but he contended that even though evidence based on actual creative performance strongly favored domain specificity (as demonstrated earlier in this chapter), one might also consider other kinds of evidence. Arguments for domain generality typically do not look at creative performances – at actual creative products or artifacts, things people
have made or written or said or done – but instead focus on psychometric and personality data.

[The conclusions of researchers using the CAT are almost always that creativity is predominantly task or content specific . . . [but] researchers utilizing traditional psychometric methods usually conclude that creativity is predominantly content general. (Plucker, 1998, p. 181)

There are essentially two kinds of psychometric evidence that have been used to look at the generality/specificity issue: divergent-thinking tests and creativity checklists. We will look at each in turn.

Plucker (1998) wrote that “researchers approaching creativity (especially divergent thinking) from a psychometric perspective over the past 50 years have worked under the assumption that creativity is content general (e.g., Guilford, 1967; Torrance, 1974),” and this assumption led to a similar assumption regarding divergent thinking. As a result, all divergent-thinking tests (including the ubiquitous TTCT) are rooted in the idea that any divergent-thinking task one might use on such a test would measure across-the-board divergent-thinking skill and that such tasks are not “specific only to the task or content area addressed in a particular divergent-thinking test” (Plucker, 1998, p. 179).

This presents a problem if one wants to use such tests to assess the domain generality or specificity of creativity for at least three reasons:

1. It’s rather difficult to trust a test that has already assumed the truth of what it is being used to prove. There is an inherent circularity with using a divergent-thinking test that was constructed under the assumption of domain generality to test whether what the test is supposed to predict – domain-general creativity – actually exists, or if creativity is instead domain specific.

2. Any tests that assume domain generality can be valid only if creativity is in fact domain general. If creativity is domain specific, then the domain-general test that is being used to see whether creativity is domain specific would necessarily be invalid. As a result, it would not be possible for a divergent-thinking test such as the TTCT to provide valid evidence for domain specificity (although, as we will see, the research using the Torrance Tests does provide evidence of domain specificity – but that evidence only calls into question the validity of the Torrance Tests).

3. If one is to use domain-general tests of creativity to demonstrate domain generality of creativity, the only kind of prediction that can be made is that the tests would predict (or be correlated with) creativity in many different domains. But what kinds of evidence might one use to assess levels of creativity across domains? The obvious answer is
creative artifacts that subjects have produced in a variety of domains. However, research has already shown rather conclusively, as Plucker (1998) himself acknowledged, that assessments of creativity based on actual creative products yield almost no evidence of domain generality.

It therefore seems unlikely that divergent-thinking tests can provide valid or convincing evidence in the generality/specificity debate, but it is nonetheless instructive to see what kinds of evidence have been produced using divergent-thinking test data. Much of this evidence resulted from efforts to validate divergent-thinking tests.

The TTCT are the most widely used divergent-thinking tests, and they come in different versions, one verbal and the other figural. Kim (2009, 2011a, 2011b), who took the pro-TTCT position in the American Psychological Association debate (as discussed in Chapter 1) on the validity and value of these tests, described the tests in this way:

TTCT-Verbal and TTCT-Figural are two versions of the TTCT. The TTCT-Verbal has two parallel forms, A and B, and consists of five activities: ask-and-guess, product improvement, unusual uses, unusual questions, and just suppose. The stimulus for each task includes a picture to which people respond in writing (Torrance, 1966, 1974). The TTCT-Figural has two parallel forms, A and B, and consists of three activities: picture construction, picture completion, and repeated figures of lines or circles. [For the TTCT-Figural], ten minutes are required to complete each activity. In Activity I, the subject constructs a picture using a pear or jelly bean shape provided on the page as a stimulus. The stimulus must be an integral part of the picture construction. Activity II requires the subject to use ten incomplete figures to make an object or picture. The last activity, Activity III, is composed of three pages of lines or circles, which the subject is to use as a part of his or her picture (Torrance, 1966, 1974, 1990, 1993; Torrance & Presbury, 1984). (Kim, 2006, p. 4)

There are various subscores to each of the Torrance Tests, but the overall creativity index score is the score most commonly reported and often the only one used, even though Torrance himself warned against this:

Torrance has discouraged the use of composite scores for the TTCT. He warned that using a single score like a composite score may be misleading because each subscale score has an independent meaning. (Kim, Cramond, & Bandalos, 2006, p. 461)

Although each of the two versions of the TTCT bears the name of the domain its tasks are derived from, both tests – figural and verbal – are most commonly used not as measures of creativity in their respective domains (and only in those domains) but rather as measures of creativity more generally (Kaufman et al., 2008b).

What would seem the most obvious first step in using these tests to assess generality/specificity would be to do what has been done with performance measures of creativity: look to see if the two measures are
correlated with one another. As already noted, Torrance himself found they were not:

Responses to the verbal and figural forms of the TTCT are not only expressed in two different modalities . . . but they are also measures of different cognitive abilities. In fact, Torrance (1990) found very little correlation ($r = .06$) between performance on the verbal and figural tests. (Cramond et al., 2005, pp. 283–284; italics added for emphasis)

These two tests are not intended to be domain-specific measures of creativity in the verbal and figural domains, however, so the parallel with measures of actual creative performance many not be appropriate. The fact that they are essentially orthogonal measures with no shared variance does raise the question of validity, however. How could two valid tests of the same construct (whether conceptualized as either general divergent-thinking skill or general creative-thinking skill) have no shared variance? This would be like discovering that two different IQ tests were totally uncorrelated with one another. One would have to assume that either one or both of the tests must be invalid. And even if one of the tests is valid while the other not, unless one knows which is the valid one, it would be hard to trust any research based on either of them.

Almost all of the validation efforts for the TTCT were longitudinal studies conducted long ago by Torrance (1969, 1972) himself. These studies suffer from serious threats to validity. As early as 1972, Crockenberg had argued that “given the creativity criteria used . . . [the results of these studies] should not be taken too seriously” (p. 35). Kogan noted in 1983 that “evidence for the long-term predictive validity of the Torrance instruments must be considered equivocal at best” (p. 650). Baer (1993) summarized research on the Torrance validation studies and concluded that the measures used to validate the tests might be valid measures of particular kinds of creativity “and yet be of little or no value in predicting other kinds of creative performance” (p. 37).

More recently Plucker (1999) completed a full reanalysis of the Torrance longitudinal data (Torrance, 1969, 1972). His results suggested that the verbal TTCT scores did predict many of the kinds of things that had been used as evidence of creative performance (these were largely things subjects had self-reported as personal accomplishments from a checklist of creative achievements; see upcoming sections for validity problems associated with self-report scales), but figural TTCT scores did not predict the same outcomes. Given Torrance’s own finding that the two tests were essentially uncorrelated, this should perhaps not come as a surprise, but this finding raises the same question that the lack of correlation between the tests raises. Based on these results, creativity and the TTCTs can not all be domain general. Nothing in these data argues against domain specificity,
However; in fact, it is exactly the kinds of results domain specificity would predict — correlations among measures of creativity on similar tasks and lack of correlations on measures of creativity in different domains.

Although he had argued for domain generality, Plucker (1999) explained these results — creative achievement, as measured by self-report questionnaires, being positively correlated with the verbal TTCT but not with the figural TTCT — in exactly the same way a domain generality theorist would explain them:

The importance of verbal DT relative to figural DT may be due to a linguistic bias in the adult creative achievement checklists. For example, if a majority of the creative achievements required a high degree of linguistic talent, as opposed to spatial talent or problem solving talents, the verbal DT tests would be expected to have a significantly higher correlation to these types of achievement than other forms of DT. (Plucker, 1999, p. 110)

Without wading further into the question of the validity of any of the Torrance Tests (which will be considered in more detail in Chapter 5 on creativity assessment), it appears that to the extent these studies of divergent-thinking tests can tell us anything about the generality/specificity question, what they have to say favors domain specificity.

The other kind of psychometric data that is sometimes used to argue for content generality is based on creativity checklists, which are a form of self-report. Plucker (1998) wrote that “performance assessments produce evidence of task specificity, and creativity checklists and other traditional assessments suggest that creativity is content general” (p. 180). Lubart and Guignard (2004) came to a similar conclusion, writing that “performance-based evaluations provide results favoring a domain-specific view, whereas self-report inventories lead to a more general-oriented conception of creativity” (p. 53). Plucker (1998) gave as an example of this a study by Runco (1987) that used students’ self-reported levels of creativity in seven performance domains:

Runco (1987) compared students’ creativity checklist responses to quality ratings of the students’ creativity (scored using a technique not unlike the CAT). The students’ checklist scores provided evidence of content generality, and the quality ratings suggested content specificity. (p. 181)

So even those arguing for domain generality seem to have conceded that studies based on actual assessments of participants’ creativity, such as the “quality ratings” that Runco (1987) accepted as heralds of domain specificity, the “performance assessments” that Plucker agreed were evidence of domain specificity, and the “performance-based evaluations” that Lubart and Guignard (2004) conceded were indicators of domain specificity. The kind of evidence that these authors offered as a way to
keep the door open to domain specificity, the final kind of evidence that might be considered, is based on self-reported creativity checklists.

Self-report scales like the one that Runco (1987) used and that Plucker (1998) and Lubart and Guignard (2004) have cited do tend to support at least a modest degree of domain generality. For example, Hocevar (1976) found “low to moderate” (p. 869) correlations (ranging from 0.17 to 0.76) among self-report indexes of creativity in various domains for college students. In a study in which several thousand subjects self-reported their own creativity in 56 domains, Kaufman, Cole, and Baer (2009c) found both an overarching general factor and seven more specific areas of creative performance. Figure 2.1 shows the seven factors and their relationship with (standardized loadings on) a general creativity factor.

Self-report data on creativity thus suggest that there may be a domain-general factor in creativity, a conclusion that conflicts with the findings of research based on actual creative performance. Self-reported creativity may be a problematic source of information, however. Hocevar (1981) claimed that such self-report scales were “perhaps the most easily defensible way to identify creative talent” (p. 455), but many others have

FIGURE 2.1 General Thematic Areas from Kaufman, Baer, and Cole (2009b).
questioned the validity of self-report scales, both in creativity research and more generally (Azar, 1997; Brown, 1989; Rowe, 1997).

College students cannot accurately estimate their own IQs (e.g., Paulhus, Lysy, & Yik, 1998), despite the extensive information they would be expected to have received in this area (e.g., SATs and other standardized test scores, high school and college grades, success in college applications, and of course actual IQ test score reports). In a study of workplace performance, DeNisi and Shaw (1977) showed that self-reported ability did not predict test scores and should not be used in place of ability tests. The validity problems that self-report data have experienced in arenas with much higher stakes than most creativity research can be both large and troubling. For example, in one medical study, “doctors self-reported their hand-washing rate at 73 percent, whereas when these same doctors were observed, their actual rate was a paltry 9 percent” (Dubner & Levitt, 2006). One might argue that it is hardly surprising that people will sometimes misrepresent themselves when doing so is likely to benefit them in some tangible way, but people have also been shown to self-report erroneously—even when they knew that their self-reports were erroneous—in astonishingly large numbers even when it can result in significant financial loss, apparently just to make themselves look better (Dubner, 2007). Even when the facts are clear to the “reporters” (the selves of self-reported data), strong evidence indicates that those self-reports may result in seriously false reports. Is it any wonder skepticism about trusting self-report data is high when it is far less likely that the reporters will actually have factual information to report but must instead base their reports on their own feelings, hunches, and beliefs about themselves?

Brown (1989) inveighed against the use of self-reported creativity in research, arguing that in assessing creativity, “self-report data and retrospective case histories are generally unverifiable” (p. 29). This inability to validate the data obtained from self-report measures makes one hesitant to rely heavily on such data, and evidence shows that such self-report data may be not just slightly but sometimes totally invalid measures of creativity. Kaufman, Evans, and Baer (2010c), for example, reported that fourth-grade students’ self-assessments of their own creativity did not match the assessments of experts in any of the four domains tested (math, science, writing, and art). It wasn’t simply that the students tended to inflate their self-assessments. Their self-assessments actually bore no relationship whatsoever to the ratings experts made of their actual creative products (although none of the correlations was statistically significant, three of the four were negative and the mean of the four correlations was –0.075). This kind of data brings to mind the famous Groucho Marx quote, “Who are you going to believe, me or your lying eyes?”

This lack of evidence of validity alone makes one hesitant to rely very heavily on such data, and other recent research suggests that self-reported
creativity does not match experts’ judgments of creativity. Dollinger, Burke, and Gump (2007) found that although the Creative Behavior Inventory showed strong reliability, it correlated only 0.16 (nonsignificant) with three rated creative products (a drawing, a story, and a photo essay) produced by college students. The limitations of self-report data extend well beyond creativity research, of course (see, e.g., Rowe, 1997).

Reiter-Palmon, Robinson, Kaufman, and Santo (2012) recently reviewed several frequently used self-report measures of creativity and described their work and their results as follows:

Self-evaluations or self-perceptions of creativity have been used in the past both as predictors of creative performance and as criteria. Four measures utilizing self-perceptions of creativity were assessed for their usefulness as criterion measures of creativity. Analyses provided evidence of domain specificity of self-perceptions. The scales correlated with self-report measures of creativity, but not with objective measures. Self-perceptions of creativity had strong to moderate relationships with personality and creative self-efficacy. These results suggest that although self-perceptions of creativity may provide some information about creativity, researchers should be cautious when using this measure as a criterion. (p. 107)

Their “[r]esults support the notion that self-perceptions of creativity are domain specific” (p. 112), but because of their caution in using these scales as criterion measures in creativity research, it seems unfair to count them as evidence of domain specificity any more than one should conclude from some other study that might report self-perceptions of creativity favoring domain generality. As they wrote, “[s]elf-perceptions of creativity did not correlate (even with the large sample size) with various measures of creative problem solving” and “the correlations with the more objective measures of creative accomplishments” (p. 112) were low, so it appeared that self-perceptions of creativity did not jibe with real-world measures of actual creative performance. If the criterion measures can’t be trusted, then no conclusions based on them should be trusted either. The problems with using self-reported creativity when looking for domain generality ran even deeper, however. “Even more problematic [than the lack of correlations with objective measures of creativity] were the findings that correlations between self-perceptions of creativity across domains and the measures of creative personality and creative self-efficacy were high” (pp. 112–113). These results would especially contaminate any research that was looking for correlations across domains that might be caused by domain generality because such results, even if found, might actually reflect subjects’ beliefs about creative personality and creative self-efficacy, not domain generality.

Silvia, Wigert, Reiter-Palmon, and Kaufman (2012) recently reviewed what they termed “four new and promising [creativity self-report] scales” (p. 19) and concluded that, although most such reviews “end on a grim
note” (p. 31), these new scales were more promising, at least if used only for low-stakes assessment. The problem of deception remains, however: “‘Faking good’ is easy to do on these scales for respondents who want to appear to be more creative than they really are, so scores from high-stakes contexts will be suspect” (p. 32). This seems particularly problematic if such scales are used to judge domain generality of creativity, because it seems likely that subjects who wish to “fake good” are likely to self-report inflated scores across domains, whereas those who don’t try to “fake good” will self-report more realistic (and thus generally lower) scores across domains. If, as Silvia and others suspect, only some subjects in low-stakes assessments will choose to “fake good,” then the result would skew cross-domain correlations toward a domain-general outcome that is really a measure of how much (or little) subjects choose to “fake good,” not how consistent their creativity was across domains. It is certainly to be hoped that these newer self-assessment tools will prove to be more valid than their predecessors, but it seems unlikely that even if they do will they be able to contribute much to the generality/specificity debate unless the “faking good” problem can be solved. Given the validity problems that continue to haunt creativity self-assessments, using them at this point as evidence to support any conclusion in the generality/specificity debate is simply courting confusion.

Self-rated reports of creativity in different domains are suspect and, at least until recently, have failed to pass even the most basic tests of validity. They remain popular because they are easy to do; one can avoid the hard work of having subjects actually create things and the expense of having groups of experts rate those products for creativity (which is the downside of the kinds of CAT-based research studies reported above – high validity, but also high costs). One way to get around the validity problem associated with creativity self-ratings without undertaking the hard work and expense of using actual creative products as one’s source of data about creativity is to study the careers of successful creators, people who have been acknowledged by their fields to be creative based on not just one performance but a lifetime of work. What one can do with such data is necessarily limited in regard to the generality/specificity question, however. Most eminent creators have become notable in just one field (and it is this judgment of their fields that validates their creativity; the fact that they may have other interests does not provide evidence of creativity in those other areas), so they cannot offer any direct evidence of domain generality. But studies of eminent creators can provide some indirect evidence, and this is the one area in which some support for domain generality can be found. (Polymaths who have had success in more than one field are a special case, which will be discussed separately in the next section.)

Studies of creative people who have achieved some level of eminence eliminate the need for such self-ratings (although studies of such creators
still typically use self-ratings of personality, where even eminent creators may not be experts, so *caveat emptor*). In these studies, the creativity of successful artists and scientists and others has been judged by their colleagues in their respective domains, and those experts have concluded that they have contributed significantly to their fields. We therefore are not dependent on how these highly creative people might self-rate their own accomplishments.

Feist (1998, 1999) looked at 50 years’ worth of this kind of research, including in his meta-analysis any studies that considered possible connections between creativity and personality. His review produced three comparisons: scientists with nonscientists, more creative versus less creative scientists, and artists versus nonartists. He found both domain-based differences and similarities across domains. Here is a summary of his conclusions regarding domain-based differences in creative personalities:

Creative people in art and science do not share the same unique personality profiles: Artists are more affective, emotionally unstable, as well as less socialized and accepting of group norms, whereas scientists are more conscientious. (Feist, 1999, p. 290)

Feist (1998, 1999) was careful to note that the observed correlations do not allow inferences regarding causation. It is possible that the personality traits that seemed to be shared by eminent creators in his two very different domains had helped produce creative behavior in both domains, but is equally likely that creative success in any domain might lead to some similar personality traits (e.g., it may be that the personality traits observed in highly creative people did not cause creativity in any domain but were themselves caused by or in some way the result of success). This idea can be easily understood by considering such traits as self-confidence, ambition, and dominance, all of which were traits shared by creative scientists and artists. One can certainly see how self-confidence, ambition, and dominance might lead to creativity, especially at the highest, Big-C level; but one could as easily see how these traits might be the result of (rather than the cause of) creativity. People who have been successful in their fields (even if their creativity is limited to a single domain) might be expected to become more self-confident, more ambitious, and more dominant. The fact that one could make sense of a causal arrow going in either direction does not allow one to know which is cause and which effect (or even if these traits co-vary with no causal connection). It is also possible that some of the traits that were shared by highly successful creators in the studies Feist reviewed might indeed influence success in their fields at the highest levels (Big-C creativity) but have less relevance to the more everyday (little-c) creativity that is the focus of most creativity research (including the CAT-based studies that form the solid core of research supporting domain specificity).
It is also not clear whether the shared personality traits that Feist (1998, 1999) uncovered are truly shared across domains. Is the observation that creative artists and scientists both tend to be open to new experiences evidence of a general trait, or a domain-specific one? Artists and scientists might be open to different kinds of experiences. If openness to experience means openness only to certain kinds of experiences (perhaps those in a particular domain or general thematic area), then that openness is domain specific and will not translate to creativity in other domains, only the one in which the individual is indeed open to new experiences. In the same way, it is not surprising to learn that more creative people may be more highly motivated (more ambitious, more driven) than their less creative counterparts, but as Mlodinow (2008) suggested, “successful people in every field are almost universally members of a certain set – the set of people who don’t give up” (p. 11). This may well result in more (or at most a very few) success (and more creative performance), but even if the causal arrow does go from motivation to creativity, this finding begs a more basic question: Is motivation a common, domain-general resource that can be deployed widely and in any domain, or is it specific to a domain? Might a creative playwright who spares no pains when it comes to improving her writing show the same resolve when it comes to playing tennis, solving math exercises, or arguing with her accountant? Recall Feist’s (2004) argument that if domain generality were true, “a creative person could be creative in any domain he or she chose” (p. 57). Was Mozart’s interest and talent in music simply one of many choices he might have made? Might he instead have chosen sculpture, chemistry, physics, or architecture as a career and had the same measure of creative success? And might his drive and ambition, his openness to new ideas, his dominance, and his self-confidence all have had a much more limited focus that did not encompass most other domains?

These caveats are important ones, but personality measures nonetheless do provide a limited kind of support for at a modest degree of domain generality, especially when considering the highest levels of creative performance (Big-C creativity). These findings stand in stark contrast to those based on assessments of actual creativity across domains that almost universally support domain specificity. It is important to remember that the scientists and artists that Feist (1998, 1999) studied did not actually provide any evidence of domain generality; they each were creative in one domain (art or science) or the other, but not both. There are reasons why highly creative people tend to be creative in only one (or at most a very few) domains, reasons that would limit the extent of their creative accomplishments even if creativity were domain general (as will be discussed in the next section about data that, although appear on the surface to be relevant to the generality/specificity debate, in fact tell us nothing because both sides make very similar predictions). But it is fair to conclude that
personality measures, although not providing clear or direct evidence of domain generality, do leave open the possibility of some personality traits that might, especially at the highest levels of performance, positively impact creative performance.

False Evidence: Why the Existence of Polymaths Cannot Help Resolve the Generality/Specificity Question

When I talk to people outside the creativity research community about the domain specificity of creativity, many ask, “What about polymaths?” If creativity is domain specific, how could one person like Leonardo da Vinci be so creative in several very different fields? Others draw the opposite conclusion: “I guess that’s why there are so few all-around geniuses,” often adding, “like da Vinci,” who seems to be everybody’s favorite polymath.

Truly extraordinary people such as da Vinci are certainly interesting, and he’s not alone, but polymaths really can’t help us with the question of the domain specificity or generality of creativity. Just as domain generality doesn’t insist that if someone is creative in one area they must be equally creative in all areas (i.e., domain generality doesn’t say that every creative genius must be a polymath, or that everyone who is mildly creative in one domain must also be mildly creative in every other domain), domain specificity doesn’t argue that people can only be creative in a single domain (i.e., domain specificity doesn’t rule out polymaths – in fact, domain specificity predicts their existence).

To help understand what predictions domain specificity would make about polymaths, it may help to consider an analogy that is unrelated to creativity. Consider ways in which height and piano-playing skill might be related. Let’s assume that height and piano-playing skill are totally unrelated – that they have a correlation of zero. The following possibilities remain:

1. A person could be tall but not be a concert pianist.
2. A person could be a concert pianist but not be tall.
3. A person could be neither tall nor a concert pianist.
4. A person could be both tall and a concert pianist.

A zero correlation just means that knowing one thing about a person tells us nothing about some other thing.

If creativity is domain specific, then the correlation between creativity in two distinctly different domains, such as dance and cosmology, would be zero. If that were the case, it would still be true that:

1. A person could be a creative dancer but not be a creative cosmologist.
2. A person could be a creative cosmologist but not be a creative dancer.
3. A person could be neither a creative dancer nor a creative cosmologist.
4. A person could be both a creative dancer and a creative cosmologist.

One could add more traits – concert pianist, being tall, having a fear of heights, and being a good cook, say, which for the sake of argument I will assume are unrelated attributes – and even if there were a zero correlation between every pair of attributes, it would still be possible for someone to be all four a tall concert pianist who is a good cook and afraid of heights. (The more unrelated traits added to such a list, the less likely it would become for anyone to possess all of those traits, but it would still be possible and, if the population were large enough, expected.) Similarly, no matter how many unrelated domains there might be (in terms of creativity), someone could be creative in 1, 2, 3, 4, or many more of those domains. In fact, other things being equal, one would expect to find a few people who were extremely creative in several domains, some who evidenced little creativity in any domain, and all the rest, would be expected to be creative in varying degrees in every possible set of mutually unrelated domains. So if creativity is domain specific, then one would expect some people to be highly creative in more than one domain, and finding such people doesn’t count as evidence against domain specificity.

Domain specificity doesn’t argue or expect that people can be creative in only a single domain. It says only that because whatever things lead to creativity in different domains are different, creativity in one domain does not predict creativity in other domains. Domain specificity argues that to the extent domain-based creativity-relevant talents are randomly distributed, one would expect a few people to be creative in many domains. The presence of a few da Vincis does not disprove domain specificity – just the opposite. It is exactly what domain specificity predicts.

Domain specificity does predict that people who are creative in a domain will be creative when performing different tasks within that domain; the more closely similar the tasks are, the higher should be the observed correlations of creativity on those tasks. And that is exactly what the research shows us.

So perhaps the scarcity of multidomain creative geniuses – Why aren’t there more Leonardo da Vinci’s? – is evidence of domain specificity. If creativity is domain general and someone has enough of those domain-general skills or traits to reach creative eminence in one domain, then shouldn’t they achieve similar levels of creativity across the board? No, one should not expect this outcome, at least not unless the domain-general skills or traits were the only things that influenced creative performance. But no one, even the most committed domain-generality theorist, thinks that domain-general skills are the only things that influence creative performance. Every domain also requires many other things (specific skills, specific knowledge, specific equipment, specific motivation, specific
talents, etc.) that promote creativity in that domain but not in others. One needs a guitar to be a creative guitarist, but not to be a creative geologist, and one needs an interest in rocks to be a creative geologist, but not to be a creative guitarist. (An interest in rock music, not rocks per se, might be relevant, of course.)

There is at least one more reason why Joni Mitchell is an extremely creative songwriter but not an extremely creative cosmologist. It generally takes many years to acquire the kinds of domain-specific skills and knowledge needed before “even the most noteworthy and ‘talented’ individuals” (Weisberg, 1999, p. 230) can make a truly original contribution to a field—something creativity researchers call the “ten-year rule” (Hayes, 1989). As Gruber and Davis (1988) wrote, “Perhaps the single most reliable finding in our studies is that creative work takes a long time” (p. 264). This extended period of intense preparation must be spent in “deliberate practice and the development of expert performance” (Weisberg, 1999, p. 233). Because it takes about 10 years, give or take, just to prepare oneself for the kind of paradigm-shifting work that may someday be recognized as a work of genius, it should come as little surprise that few people manage to reach the highest levels of creative accomplishment in a dozen or more fields in a single lifetime.6

In his analysis of the kinds of “mechanisms that mediate superior performance,” Ericsson (2003) found the underlying abilities that led to performance at the highest levels were “surprisingly complex mechanisms highly specific to the task domain” (p. 109). This finding is consistent with the ten-year rule’s argument that years of intense domain-specific study and practice are needed before anyone can make a Big-C-level creative contribution to a domain. (Perhaps not coincidentally, the training time for high-level sports performance is similar; Syed, 2010.) It is somewhat ironic that it is the need for years of domain-specific preparatory work that shields domain generality from the need to explain why there are not considerably more polymaths, as domain generality would otherwise necessarily predict.

So the existence of polymaths cannot tell us much about the domain generality or domain specificity of creativity (Kaufman, Beghetto, &

---

6The example of Joni Mitchell is an interesting one because she has made something of a mark in more than one domain. She is also a successful painter (and if she had set her sights on cosmology, who knows?). It’s certainly true that many highly creative people have other creative interests outside the field in which they have become famous, as Root-Bernstein and Root-Bernstein (2004) noted. Georgia O’Keeffe once told Joni Mitchell, “I would have liked to have been a painter and a musician, but you can’t do both.” Mitchell replied, “Oh, yes, you can!” (Weller, 2008, p. 427), based perhaps on her own prodigious output as a musician and as a painter. But being highly creative in just two domains is indeed extraordinary, and it is almost impossible to find an example of anyone today being creative at the highest level in three or more domains.
Baer, 2010a; Kaufman, Beghetto, Baer, & Ivcevic, 2010b). Some writers have indeed seemed to argue that the existence of polymaths is evidence for domain generality (e.g., Root-Bernstein & Root-Bernstein, 2004), but domain specificity predicts occasional polymaths, so pointing out their existence isn’t really an argument against domain specificity; as interesting as polymaths may be (and they certainly are interesting), as far as the generality/specificity issue is concerned they are nothing but a sideshow that diverts attention from data that actually can speak to the question, data that offer little room for domain generality. Domain generality also predicts polymaths, of course; in fact, it predicts many more polymaths than we actually observe. If creativity were domain general, then we should have large numbers of polymaths because if one has a lot of domain-general creative ability, it should lead to high levels of creativity across the board (just as people with a lot of g tend to get high marks in lots of areas). Domain generality is shielded from needing to defend this failed prediction by the ten-year rule (Hayes, 1989), however, because no one has time to put ten or more years of intense work into several domains. So the existence of polymaths really prove nothing either way.

Simonton (2006) challenged the ten-year rule, at least in its strictest interpretation, by showing that the greatest geniuses typically spend fewer years than their less-talented peers in knowledge acquisition in their domain before exhibiting their remarkable creativity. But he also has shown that the greatest geniuses produce the greatest quantity of work. This is what he calls the “equal-odds rule,” which argues that “quality should be a probabilistic consequence of quantity” (p. 54) – which also limits the likelihood of creative genius in multiple fields. There just may not be enough time to produce a sufficiently large quantity of works in multiple domains.

Polymaths are fascinating, certainly, and they show that it is possible to be creative – sometimes even at the highest levels, although this is exceedingly rare – in multiple domains. But that is exactly what both domain generality and domain specificity theories predict. Both the single- and the multiple-domain talents of the most highly creative people – geniuses at the Big-C level of creativity – simply can’t tell us much, one way or the other, about the domain generality/specificity question.

What Can We Conclude?

In his textbook Explaining Creativity, Sawyer (2012) concluded that “[a] wide range of studies has shown that much of creative ability is domain-specific” (p. 60). The best kind of evidence – evidence based on actual creative performance – is clear in its support of the domain specificity of creativity. Other kinds of evidence do little to contest this claim. Although self-reports of creativity in different domains suggest a place for domain
generality, self-reported creativity data have so little validity that they will convince only those already convinced. The existence of polymaths tells us nothing about generality/specificity because both theories make essentially the same (correct) prediction. And personality studies of eminent creators are interesting, but can at most suggest, very indirectly, that there may be some traits shared by highly creative people in different domains – and even this tentative finding cannot be generalized beyond those working at the highest (Big-C) levels of creativity.

For many decades, going back even beyond Guilford’s 1950 American Psychological Association Presidential Address, creativity was assumed – to the extent that creativity was studied at all by psychologists – to be rooted in a domain-general set of skills and traits. That assumption went unchallenged for many years, but in the past quarter century the tide has turned, as even domain-generality theorists like Plucker (1998) agree, and domain specificity theory is in ascendance. Beyond the contributions of general intelligence, the impact of domain generality can only be slight. The following four chapters explore what that means for creativity theory, research, testing, and teaching.