An adult observed a preschool child working assiduously on a drawing and asked, “What are you drawing?” Without looking up, the child responded, “I’m drawing the face of God.” The adult smiled and said, “But no one knows what God looks like.” The child answered, “They will in a couple minutes.”

No one is really sure what creativity looks like either, but half a century ago Joy Guilford described one important piece of it—divergent production—and Guilford’s vision has shaped what many of us think creativity looks like, especially in the area of creativity assessment.

The Torrance Tests are essentially divergent thinking tests based on Guilford’s model, and they are very widely used as creativity tests. Although the Torrance Tests have lost some of the edge they had quarter of a century ago, when Torrance and Presbury (1984) reported that the Torrance Tests had been used in three quarters of all recently published studies of creativity, they nonetheless remain an important force, although perhaps more in schools than in research. If they are valid measures of creativity, that’s wonderful. If they are not, it’s a huge problem. It would mean that we may think we know a lot of things—things we’ve learned from research by using a possibly invalid tool—that we don’t really know.

Guilford grouped his divergent-production factors into four categories, and those four categories—fluency, flexibility, originality, and elaboration—were the basis of the Torrance Tests for many years. Torrance cautioned against the use of a composite score, recommending instead that the subscales be interpreted “in relation to one another” to get a picture of an individual’s skills:

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, Crammond, & Bandalos, 2006, p. 461)

But this has not been how the tests have generally been interpreted. Contrary to Torrance’s own advice, composite scores are widely used often with no mention whatsoever of subscale scores.

The Torrance Tests and their scoring systems have been revised over the years—they are now a half century old—but they remain, despite their name, essentially measures of divergent thinking, a theorized component of creativity. There have long been two forms of the tests, figural and verbal, which Torrance reported in 1990 had almost no relationship to one another, showing a correlation of only .06:

Reponses 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. (Crammond, Matthews-Morgan, Bandalos, & Zuo, 2005, pp. 283–284)

Therefore, are the Torrance Tests best thought of as measures of particular aspects of divergent thinking that need to be interpreted in relation to one another, as Torrance suggested, and not as a composite, overall creativity measure? Are the Torrance tests measuring divergent-thinking ability in two discrete and totally unrelated domains, as the lack of relationship between the figural and verbal forms suggests? The answer, based on how the tests are actually used, seems to be neither. Subscale scores that measure different aspects of divergent thinking are routinely ignored in favor of overall creativity index scores, especially by gifted-talented programs, which are the most active users of the Torrance Tests (Scholastic Testing Service, 2009). In addition, researchers now often argue that the overall Creativity Index is the best predictor of creative ability (e.g., Plucker, 1999; Yamada & Tam, 1996). One has to wonder how we’ve gotten to a place where these tests, despite the arguments of Torrance against the use of composite scores and the evidence that he himself produced suggesting that his tests measure creativity in particular domains, are primarily used as measures of creativity in the most general sense.

I believe the fact that Torrance himself gave us sound reasons to question the widespread current use of his tests as general measures of creativity should make even the most ardent advocate of these tests somewhat skeptical, but I think the problems run even deeper than this. My argument focuses on two key issues:

1. The evidence that the Torrance Tests measure anything significantly related to creativity is open to question. The main validity studies supporting use of the tests has been called into question repeatedly, and there is counterevidence that challenges the tests’ validity.

2. The Torrance Tests have had the unfortunate (albeit unintended) effect of distorting how people in and out of the field of creativity research think about creativity.

For these reasons, we would be better off setting these tests aside and measuring creativity in other ways.
There are very different ways to read the validity studies that have been conducted by using divergent thinking tests. Here is how two reviewers summed up the evidence: (a) “Little if any of that systematic variation [in adult creative achievement] is captured by individual differences on ideational fluency tests” (Wallich, 1976, p. 60). (b) “They appear to lack convergent and discriminant validity. Even worse, they seem to lack predictive validity as well because the correlations between test performance and overt creative behavior tend to be relatively modest.” (Simon-ton, 2007, p. 353).

Much of the work on validation stems from Torrance’s own longitudinal studies, which go back 50 years. Many reviewers have commented on these studies, some concluding that they don’t tell us what they claim to tell us and provide little in the way of validation, and others coming to the opposite conclusion. It’s generally acknowledged Torrance’s studies constitute the primary validation data for the Torrance Tests: (a) “Longitudinal studies conducted with the students who had been in two elementary schools and a high school in Minneapolis in the late 1950s form the basis of the predictive validity evidence for the TTCT” (Crammond, Matthews-Morgan, Bandalos, & Zuo, 2005, p. 285). (b) “Data that were collected from two elementary schools and a high school provide the major body of longitudinal research on the TTCT” (Kim, 2006, p. 6).

One big problem with those studies is that the outcome measures—the achievements that the Torrance scores are supposedly predicting—are self-report measures, and many of them involve making lists, which is just what many Torrance Test items ask one to do.

Are these self-reported lists of aspirations and accomplishments good measures of how creative someone is? Like many others, I’m skeptical. There is not room here to describe these measures in detail, but here is how this data was summarized by a researcher who has argued that these data are valid measures of individual creativity: “…information about the subject’s most creative achievement aspirations, and a checklist of creative accomplishments, which included type and degree of achievement in the arts, research, work innovations, inventions, and change in life philosophy” (Crammond et al., 2005, p. 285).

All self-report measures are somewhat suspect and perhaps especially so in the area of creativity. Kaufman, Evans, and I (2010) did a study recently in which we asked fourth graders how creative they were in several areas, and we also asked them to create things in those domains. It is interesting to note that they didn’t see their creativity as monolithic—they reported very different levels of creativity at different kinds of tasks—but their self-assessments and the assessments of experts of their actual creative performance didn’t match at all. They were quite poor judges of their own creativity. We titled the article “The American Idol Effect,” in recognition of the apparently poor self-assessments many contestants on that show have made of their abilities. I won’t argue that American Idol contestants are a representative sample—let’s hope not!—but each of us have had enough experience with people who misjudge their own abilities to suggest that the poor self-assessments of creativity made by fourth graders may not be due solely to age. Such self-assessments need to be taken with very large grains of salt.

In addition to the general problem of self-report scales’ validity, the particular outcome measures that are being used for validation of the Torrance Tests may have special problems. They may simply be measures of how much someone can say, or is willing to say, in response to an examiner. That is, they may tell us something about someone’s response style—these kinds of questionnaires may essentially be tapping into and giving high scores to subjects with a very fluent response style—but not provide a very good estimate of what someone has actually done or how creative that person has actually been. It is unfortunate that how much someone can or is willing to say to an examiner is exactly what the Torrance Test may be measuring. In fact, this kind of fluency—being able or willing to say a lot—tends to overwhelm Torrance Test scoring. (Some, like Kogan, 1983, have argued that fluency so overwhels the rest of the scoring that the other subscales can safely be ignored.) Therefore, the Torrance Test may be measuring a kind of fluency, and the follow-up questions may be measuring the same kind of fluency, but that may not be the same as measuring creativity either by way of the test or by way of asking people to list their accomplishments.

I would be a lot more easily convinced if what was being tested and what was being used as outcome data weren’t so much alike. It just seems that if the Torrance Tests were actually good at predicting creativity, we wouldn’t need to rely on questionable self-report measures. It would be rather easy to give the tests to people about whose creativity we happen to know something and look there for correlations, the way people often look for correlations between actual achievements and IQ test scores.

There have been some small studies, such as one I did a number of years ago, that look at divergent thinking test scores and actual creative performance (Baer, 1993). My focus was not divergent thinking testing, but I included a Torrance Unusual Uses item in a Study I was doing of middle school students’ creativity, where actual creative performance in a number of different areas was measured by using the consensual assessment technique (Amabile, 1996). These assessments of actual creative products were made by experts in the relevant domains, and as almost always happens with the consensual assessment technique, their independent ratings of creativity resulted in high interrater reliabilities. I did not find much of a connection between how well my subjects did on the Unusual Uses task and their actual creative performance on several other tasks: poetry writing (.08), story writing (.34), word-problem creation (−.19), equation creation (.09; 8th-grade students, N = 50).

I don’t claim this finding tells us a lot. It was a small study, done for a different purpose. Han (2003) did a similar study more recently with 109 subjects and different tasks. Han was also looking at the domain specificity question, and like me she found no support for domain generality. She gave her subjects two different divergent thinking tests, however, and her results were similar to mine. No combination of the six divergent thinking subtests explained any combination of the three performance-based assessments she conducted. The divergent thinking test scores and the students’ actual creative performance in the three domains were completely unrelated to one another. The divergent thinking tests did not predict actual creative performance in any of the three domains she assessed nor did any combination of the divergent thinking subtests.

This was also a relatively small study with just over 100 subjects, however. To settle the question of the Torrance Tests’ validity, which has been actively disputed for at least 4 decades,
what we need are a few large and carefully controlled studies looking at (a) actual creative performance in diverse domains and (b) performance on the Torrance Tests. Studies like that could be very informative, whereas further analyses of self-reported creativity by the same three sets of subjects (or even new sets of subjects) cannot convince anyone who has a healthy skepticism in regard to the validity of self-report data. Unless the Torrance Tests can be shown to relate to actual creative performance, the rather large cloud that hangs over their validity claims cannot be dispelled. The limited data of this kind that is available suggests that scores on divergent thinking tests do not correlate with actual creative performance, but larger and potentially more conclusive studies remain to be conducted.

This leaves me with the conclusion, along with the Torrance Tests’ many other critics, that what is usually offered as validation evidence—primarily longitudinal studies of the same three groups of subjects with outcome measures that have been widely challenged—is unconvincing. I don’t claim that the problems with the longitudinal studies, or the counter evidence that the Torrance Tests don’t predict real-world creativity very well, represent irrefutable proof that the tests are invalid. I think we just don’t know, and therefore using the Torrance Tests requires too much of a leap of faith. In addition, it might seriously distort how we understand creativity research. Here’s an example. Mayer and Dow (2004) recently studied creative problem solving and concluded that general measures (like the Torrance Tests) can lead to false outcomes. They wrote,

Training of creative problem solving has a somewhat disappointing history, because learning to solve one kind of problem rarely supports solving other types of problems. . . . [I]nsight problems are not a unitary concept but rather should be thought of as a collection of distinct types of problems. (2004, p. 397)

Mayer and Dow are saying that if our theory and our measures of creativity are invalid, we can’t get meaningful results, and that all-purpose general measures of creativity like the Torrance Tests can’t be valid because that isn’t the way creativity works. It’s based on the wrong picture of creativity.

Mayer and Dow showed how a general measure of creativity might lead us to miss actual positive results of creativity training—a false negative. Here’s another recent example of ways that Torrance Test scores may mislead us, in this case leading to a possible false positive outcome. Psychological Science recently published a report by Kéri (2009) with the provocative title, “Genes for Psychosis and Creativity.” This study reported a correlation between Torrance Test scores and a gene that has been associated with an increased risk of psychosis” (p. 1070). In fact, three of the four measures of creativity used in this study were the originality, flexibility, and fluency scores on one Torrance Test subscale—the “Just Suppose” subtest. The author described these assessments as “a classic laboratory measure” (Kéri, 2009, p. 1070), and based on subjects’ responses to this one Torrance Test subtest, the authors have concluded that creativity (in the most general sense—creativity in science, creativity in the arts, creativity in teaching, creativity in cooking, etc.) is linked to presence a particular gene that is associated with psychosis.

I won’t attempt to review here research in the contentious area of possible links between creativity and mental illness. Suffice it to say that whatever such linkages there may be, they appear to vary greatly across domains (Kaufman & Baer, 2002). However, because the Torrance Test is a “classic laboratory measure,” a researcher from another field has felt comfortable using one of its subscales as a primary measure of creativity; made the further assumption (an assumption that the Torrance Tests of Creative Thinking appear to make, despite Torrance’s own very clear evidence that even his two tests of divergent thinking were uncorrelated) that creativity is domain general; and reached a conclusion that creativity, in the most general sense, is related to a specific gene that is, in turn, associated with psychosis. I suspect that the author of this study, who is not a creativity researcher, may be unaware of the controversies surrounding the Torrance Tests, and of the tenuousness of his conclusions and the giant leaps of faith required to reach them. By claiming to be domain-general measures of creativity, the Torrance Tests promote such loose thinking and problematic conclusions. It should be noted that in this case the Torrance Test gave a positive result, and therefore the study was published. One must wonder how many studies have gone unpublished because they used the Torrance Tests and found no such correlations, even though there might well have been linkages to creativity in one domain or another, as suggested by Mayer and Dow’s research.

Here’s the bottom line on the validity controversy: If divergent thinking tests like the Torrance are not valid measures of creativity—a claim many researchers have made over the past 40 years—then these tests may have been warping how we understand and interpret a lot of creativity research. The amount of misinformation they may have generated would be staggering.

I’d like at this point to set aside the question of validation, which after a half century remains, at best, uncertain, and look at some other issues that might give us pause in using the Torrance Tests as general-purpose measures of creativity. First, let’s look at the issue of the domain specificity of creativity.

Many creativity theorists have, implicitly although perhaps not intentionally, used the g of general intelligence as their model. According to most of its adherents, g is very domain general (although they may disagree on why that is so; see e.g., Garlick, 2002; Jensen, 1998; Neisser et al., 1996; Van Der Maas et al., 2006). I think a better model would be that of expertise, which is very domain specific. No one is an all-purpose expert (although you may know people who seem to think they are experts on everything). Expertise varies by domain, and a person may have real expertise in several domains, expertise in just one or a very few domains, or no significant expertise at all. I believe the evidence shows that this is also true of creativity. It is highly domain specific, which means that a general test of creativity doesn’t really make sense, any more than a test of all-around, multipurpose, domain-general expertise would make sense.

I have presented the evidence for domain specificity many places already (e.g., Baer, 1993; Baer, 1994; Baer, 1998; Baer, 2010; Baer & Kaufman, 2005; Kaufman & Baer, 2005) and won’t do so in detail here, but the thrust of it is that if you measure actual creative performances of groups of people across multiple domains, you tend to find correlations that hover around zero. This doesn’t mean that there aren’t some people who are creative in many domains—and in fact domain specificity predicts a certain level of polymathy (Kaufman, Beghetto, & Baer, 2010). It simply means that creativity in one domain is not predictive of creativity in other domains. It is interesting to note that this is exactly what
Torrance found when he measured divergent thinking in the two
domains for which he developed tests. Being creative when writing
poetry doesn't seem to help one be a more creative chemist, and
one can’t take one’s creativity in software design and put it to use
dancing creatively or finding more creative ways to paint. Diver-
gent thinking may be important, but we may need multiple mea-
sures of it, domain by domain, for it to be useful: “Generalized
tests do not have as much predictive validity as tests more specif-
ically tailored to a particular domain... Hence, tests of divergent
thinking must be tailored to each domain” (Simonton, 1999, p. 86).

My concern here is twofold. One is that if there is not one
divergent thinking skill but many, then even if the Torrance Tests
are measuring one—or even some—of those divergent thinking
skills well, they may not be measuring the ones of interest to those
administering the test. In Plucker’s (1999) recent reanalysis of the
Torrance longitudinal data, he found that verbal divergent thinking
was a powerful predictor of the kinds of things he was looking at,
but figural divergent thinking was not:

“The importance of verbal DT relative to figural DT may be due to a
linguistic bias in the adult creative achievement checklists. For ex-
ample, 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)

This finding is in line with evidence cited already that figural
and verbal divergent thinking scores are not correlated, but what
does it mean for interpretation of the Torrance Tests? They are
typically used—both the verbal and the figural tests—as general
measures of creative potential. But because they are almost or-
thogonal measures that can hardly be the case. If one had two IQ
tests that were completely uncorrelated, one wouldn’t argue that
both could be measuring the same construct of general intelli-
gence. Ditto for divergent thinking tests that are uncorrelated.
They can’t both be measuring the same thing if they yield totally
different scores.

In addition, that brings me to a related concern. I fear that the
Torrance Tests—which unabashedly and without qualification call
themselves “Tests of Creative Thinking”—help warp ideas about
creativity, especially the ideas of educators, who are the people
most likely to encounter the Torrance Tests. The Torrance Tests
suggest two things that I think are dangerous ideas:

1. Creativity is all about coming up with wild ideas.

2. Creativity is one thing, so if you’re not creative in one
area, you probably won’t be in other areas.

Let me close by summing up my basic argument. I believe
available evidence suggests that the Torrance Tests may, at best,
be measuring divergent thinking ability in certain narrow domains,
but people are interpreting Torrance Test scores as measures of
creativity more generally, and the validation evidence just doesn’t
support such interpretations. This means that the ways the Tor-
rance Tests are being used causes false research outcomes and
unreliable and invalid decisions in such arenas as admission to
gifted–talented programs. And I am also arguing that using an
“unusual uses of a tin can” kind of test as our measure of creativity
may be warping a lot of people’s ideas about creativity in ways
that I think are potentially harmful. I fear that the equation of
creativity with wild and crazy ideas—or with long lists of such
ideas—tends to lead us away from a mature understanding of
creativity. And those things are problems even if the Torrance tests
were valid measures of creativity, which I think there’s just too
much reason and evidence to doubt. That’s why we’d be better off
without them.1

References

Amabile, T. M. (1996). Creativity in context: Update to the social psy-
Baer, J. (1993). Creativity and divergent thinking: A task-specific ap-
Baer, J. (1994). Divergent thinking is not a general trait: A multi-domain
training experiment. Creativity Research Journal, 7, 35–46. doi:
10.1080/10400419409534507
Sternberg (Eds.), Cambridge handbook of creativity. New York, NY:
Cambridge University Press.
amusement park theoretical (APT) model of creativity. Roeper Review,
27, 158–163. doi:10.1080/02783190509554310
report on the 40-year follow-up of the Torrance Tests of Creative
001698620504900402
Garlick, D. (2002). Understanding the nature of the general factor of
intelligence: The role of individual differences in neural plasticity as an
explanatory mechanism. Psychological Review, 109, 116–136. doi:
10.1037/0033-295X.109.1.116
Han, K.-S. (2003). Domain specificity of creativity in young children How
quantitative and qualitative data support it. Journal of Creative Behav-
ior, 37, 117–142.
Kaufman, J. C., & Baer, J. (2002). I bask in dreams of suicide: Mental
doi:10.1037/1089-2680.6.3.271
Kaufman, J. C., & Baer, J. (Eds.). (2005). Creativity across domains:

1 There is a rather special issue related to the use of the Torrance Tests
in the selection of students for gifted–talented programs that was discussed
during the question-and-answer session that followed the American Psy-
chological Association debate. It was suggested that divergent thinking
tests increase the diversity of students selected for such programs. This
may be true, but there is a rather significant fly in this particular ointment.
It has been argued that wealthy students get an unfair advantage on the
SATs because they can afford expensive test preparation services. Al-
though there is much dispute about how effective such services are, there
can be little doubt how easily divergent thinking scores can be raised with
very minimal training. As even one supporter of the Torrance Tests wrote,
“The scores may be susceptible to various coaching and administration
issues” (Plucker, 1999, p. 104).

If a 5-min familiarity with the general scoring rules for an IQ or SAT test
would allow a student to double her score on the test, the test wouldn’t be
used, no matter how valid it might be apart from such coaching. But it’s
easy to have that kind of impact on Torrance Test scores, at least the verbal
test. So even if the Torrance Tests were valid measures of all-purpose
creativity, that they are so easily coachable should give us pause before
using them in school settings where the scores might in any way matter.


Received July 29, 2010
Revision received July 29, 2010
Accepted August 3, 2010

---

Showcase your work in APA’s newest database.

PsycTESTS*

Make your tests available to other researchers and students; get wider recognition for your work.

“PsycTESTS is going to be an outstanding resource for psychology,” said Ronald F. Levant, PhD. “I was among the first to provide some of my tests and was happy to do so. They will be available for others to use—and will help me of the administrative tasks of providing them to individuals.”


**Questions?** Call 1-800-374-2722 or write to tests@apa.org.

Not since PsycARTICLES has a database been so eagerly anticipated!