JCB

Publication of the Creative Education Foundation ®

TONY MCCAFFREY

A Visual Representation to Quantitate, Diagnose, and Improve Creativity in Insight Problem Solving

ABSTRACT

A new visual representation for insight problems permits 22 new quantitative measures; which leads to a detailed diagnosis of a person's (or team's) creative weaknesses; which then leads to prescribing targeted, effective counter-techniques for each weakness. Currently, only two measures are consistently used for insight problem solving: the number of problems solved and the time to solve the problems. These coarse measurements do not reveal the intricate dynamics of solving insight problems. Furthermore, four commonly used creativity measures (i.e., *fluency*, *originality*, *flexibility*, and *elaboration*) are often not applied to insight problems. This new visualization permits the easy application of all four creativity measures. I challenge creativity researchers to help determine which of the 22 proposed quantitative measures are the most diagnostic for insight problem solving in isolation and, in a weighted linear combination, which might yield an effective quotient (i.e., overall measure) of insight problem solving ability.

Keywords: creativity, problem solving, innovation.

INTRODUCTION TO CREATIVITY MEASURES

Performance on insight problems, including Duncker's (1945) famous candle problem, is usually measured as the total number of problems solved from a set of insight problems (e.g., five of eight solved). Each problem is counted as either solved or unsolved. Often, the time to solve problems is also measured. These coarse measurements reveal little about which aspects of the problem that solvers find difficult. Talking aloud while problem solving (i.e., verbal protocol: Ericsson & Simon, 1993) allows experimenters to test particular hypotheses related to which aspects of the problem were noticed, which go unnoticed, which aspects of the problem were easy, and which aspects were difficult. Use of the verbal protocol, however, has not yet yielded a fine-grained, standardized, quantifiable set of measures to reveal the inner workings of insight problem solving.

In this paper, I propose a new graphical framework for recording insight problem solving behavior, which permits 22 quantitative metrics for measuring diverse aspects of solving insight problems. At this point, the graphical framework is theoretical and I call on creativity researchers to help determine which of the 22 metrics are the most informative for measuring insight problem solving performance and most diagnostic for improving performance.

 The Journal of Creative Behavior, Vol. 0, Iss. 0, pp. 1–17 © DOI: 10.1002/jocb.132

Traditionally, four metrics have been used to measure creativity: *fluency*, *originality*, *flexibility*, and *elaboration* (Guilford, 1967; Torrance, 1974, 1990a,b). *Fluency* refers to the number of responses given. *Originality* refers to the statistical infrequency of the item produced or described; and is the most widely recognized aspect of creativity (Runco & Pritzker, 1999). *Flexibility* refers to the number of distinct categories covered by a set of responses. *Elaboration* refers to the amount of detail within a particular response or set. These four measures have been applied to creativity tests (e.g., Guilford, 1967; Torrance, 1974, 1990a,b; Wallach & Kogan, 1965), but are often not applied to insight problems. The graphing method developed in the next section permits the easy and straightforward application of these four standard creativity measures to insight problems.

VISUALIZING INSIGHT PROBLEM SOLVING

Problem solving activity is generally described spatially as a bi-directional process between top-down problem framing and bottom-up problem solving (Rittel & Webber, 1984; Simon, 1995). Problem framing relates to specifying the goal with greater and greater precision. Problem solving relates to using the available resources to produce effects that move toward the goal's achievement. Current visualization methods, however, generally represent problem solving activity as a uni-directional process (Greene, Thomsen, & Michelucci, 2011; Nielsen, 2012). Figure 1 shows a bi-directional graphing method. The goal is placed at the top and the refinement of the goal into sub-goals grows downward. Resources are placed across the bottom and broken into features that grow upward. The resources are interacted together in an upward fashion as the interactions satisfy a sequence of sub-goals that ultimately satisfies the original goal at the top.

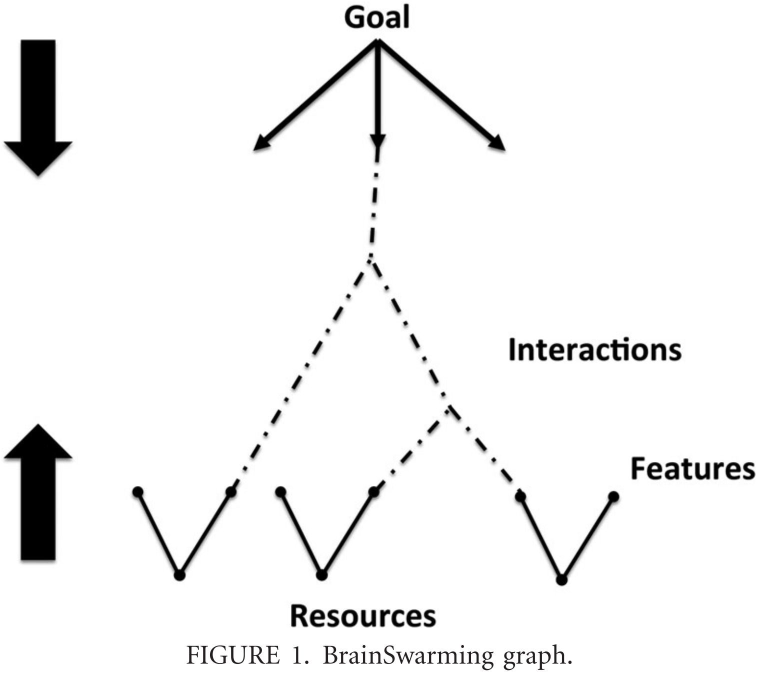

FIGURE 1. BrainSwarming graph.

When the two directions connect, a solution path emerges. Any contribution to the graph is of one of four types: *goal/sub-goal*, *interaction*, *feature*, or *resource*.

Originally called a bi-directional network (i.e., Bi-Net) in the literature (McCaffrey, Krishnamurty, & Lin, 2014), the graph in Figure 1 is also called a BrainSwarming graph (Harvard Business Review, 2014) and when used by a group provides an alternative to face-to-face brainstorming (Osborn, 1953). During group ideation, the highly visual nature of BrainSwarming permits groups to work together in parallel by simultaneously writing on a structured graph and reading each others' contributions (McCaffrey et al., 2014). In contrast, face-to-face brainstorming requires that one person contribute (i.e., talk) at a time, which slows down the ideation process when compared to simultaneous contributions. As a result of the parallel activity, no one can dominate a BrainSwarming session, as is often the case in face-to-face brainstorming unless the facilitator does a good job of controlling the dynamics. My preliminary pilot study suggests that BrainSwarming generates significantly more ideas than face-to-face brainstorming and in much less time. No study yet has compared BrainSwarming with electronic brainstorming, which consistently produces more ideas than face-to-face brainstorming (Gallupe & Cooper, 1993; Gallupe et al., 1992; Mongeau & Morr, 1999). In electronic brainstorming, people from remote locations simultaneously enter their ideas into their electronic devices and have them distributed to the other participants. In both BrainSwarming and electronic brainstorming, people can work simultaneously and in silence. BrainSwarming can be adapted to an electronic format, so the main difference is that BrainSwarming produces a structured graph of ideas while electronic brainstorming produces an unstructured list.

In this paper, however, I am interested in presenting the BrainSwarming graph as a source of new metrics for insight problem solving, so I will not engage in a more detailed comparison between BrainSwarming and forms of brainstorming.

In the next section, a sample problem will be presented so that readers can gain familiarity with how the graph grows in a structured manner. Following the example, new metrics will be presented as possible measures for insight problem solving behavior.

SAMPLE INSIGHT PROBLEM: *THE STUCK TRUCK*

Consider the *Stuck Truck* insight problem. A truck driver was driving a delivery truck under an underpass when suddenly it came to a screeching halt. The truck top wedged so tightly against the underpass that the truck could not go forward or backward. Without increasing damage to either the top of the truck or the underpass, how can the driver get the truck unstuck without help from another person?

The initial graph set-up in Figure 2 places the goal, *liberate truck from underpass*, at the top of the graph and the known resources (i.e., *truck*, *road*, and *underpass*) across the bottom. Figure 3 shows the two directions growing toward each other. Several sub-goals have been placed beneath the main goal, each of which could lead to accomplishing the main goal. Some features reach upward from the resources and any of these features might become relevant to a solution.

Figure 4 shows three candidate solutions, as represented by three places where the graph's top and bottom have become connected. First, oil could be extracted from the engine to lubricate the top of the truck sufficiently so the truck might slide free of the underpass. Second, partially deflating the tires could lower the truck sufficiently so that

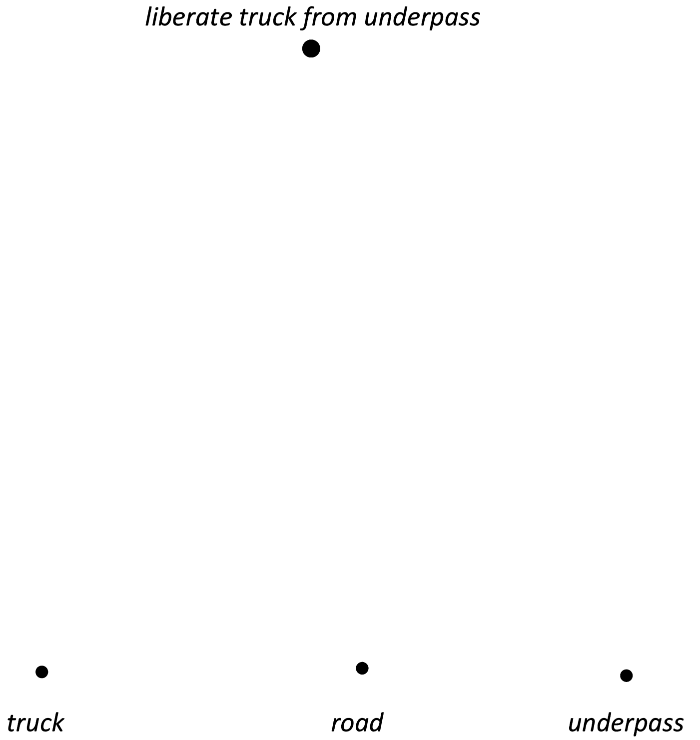

FIGURE 2. Initial BrainSwarming graph for stuck truck problem.

the truck would no longer touch the underpass. Third, the suspension and shock absorbers could be stressed by placing heavy debris (e.g., rocks) from the side of the road into the cargo bay causing the truck to lower a few inches.

NEW INSIGHT PROBLEM SOLVING METRICS

As shown in Table 1, crossing the four traditional creativity measures (i.e., *fluency*, *originality*, *flexibility*, and *elaboration*) with the five main components of a BrainSwarming graph (i.e., *goal/sub-goals*, *interactions*, *features*, *resources*, and *solution paths*) yields 20 possible metrics. Of these 20 possibilities, 18 of them result in solid candidates for metrics. A “Yes” in a cell of Table 1 indicates a new metric. Comments are placed in a cell of Table 1 when a proposed metric is not straightforward and requires a more nuanced interpretation.

As *fluency* involves counting the items in a BrainSwarming graph, it is the most straightforward metric to apply across the five BrainSwarming components. Regarding *sub-goals*, count the number of nodes in the downward growing goal tree in one of two ways. For example, in Figure 4, there are five sub-goals if one counts only the leaf nodes of the downward growing goal tree. Six sub-goals are counted if one includes the non-leaf nodes of the goal tree. Figure 4 shows three *interactions* and also three *solution paths*. Figure 4 shows three *resources*. *Features* could be measured in various ways: including

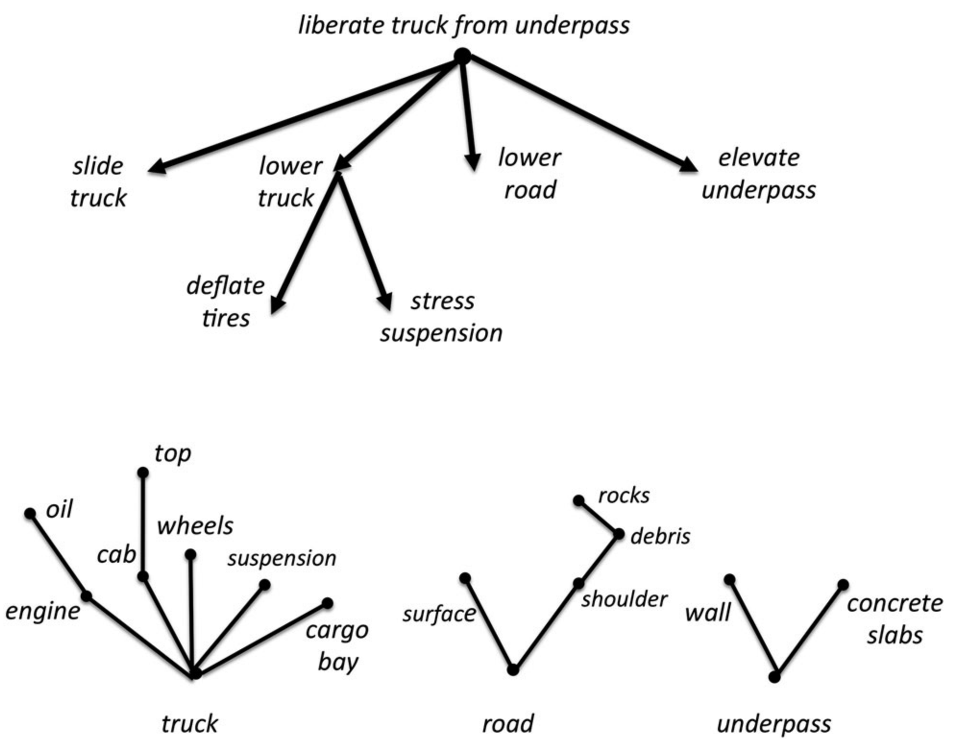

FIGURE 3. Addition of sub-goals and features.

total features for all resources, average number of features per resource, maximum number of features for any of the resources, or weighting the features deeper in the feature tree more than features at more shallow levels. In Figure 4, the truck is broken down into five features at its first level (i.e., *engine*, *cab*, *wheels*, *suspension*, and *cargo bay*) and two more features at the second level of the feature tree (i.e., *oil* and *top*). Giving the same weight to all features of the truck would yield the following score: 7 features × 1 weight unit = a score of 7. Weighting the levels differently would yield a different score: (5 features at first level × 1 weight unit) + (2 features at second level × 2 weight units) = 5 + 4 = a score of 9. The *fluency* of a resource's features is intimately related to the measure of how *elaborately* (i.e., amount of detail used) a resource is described. For example, the truck's features are more elaborately described than the underpass' features. The *fluency* measure described above permits the difference in *elaboration* between the truck and the underpass to be quantified: either 7 vs. 2 or 9 vs. 2, depending on the weighting schemed used.

Originality can be easily quantified for any graph element, but objective measures, as opposed to subjective measures (Silvia, Martin, & Nusbaum, 2009; Silvia et al., 2008), may not be the best metric for originality. Several objective measures are possible. First, count how frequently each item is mentioned among a group of problem solvers and divide the frequency count by the number of problem solvers. Second, a popular method assembles all items mentioned by the group into a pool and awards a 1 to any item

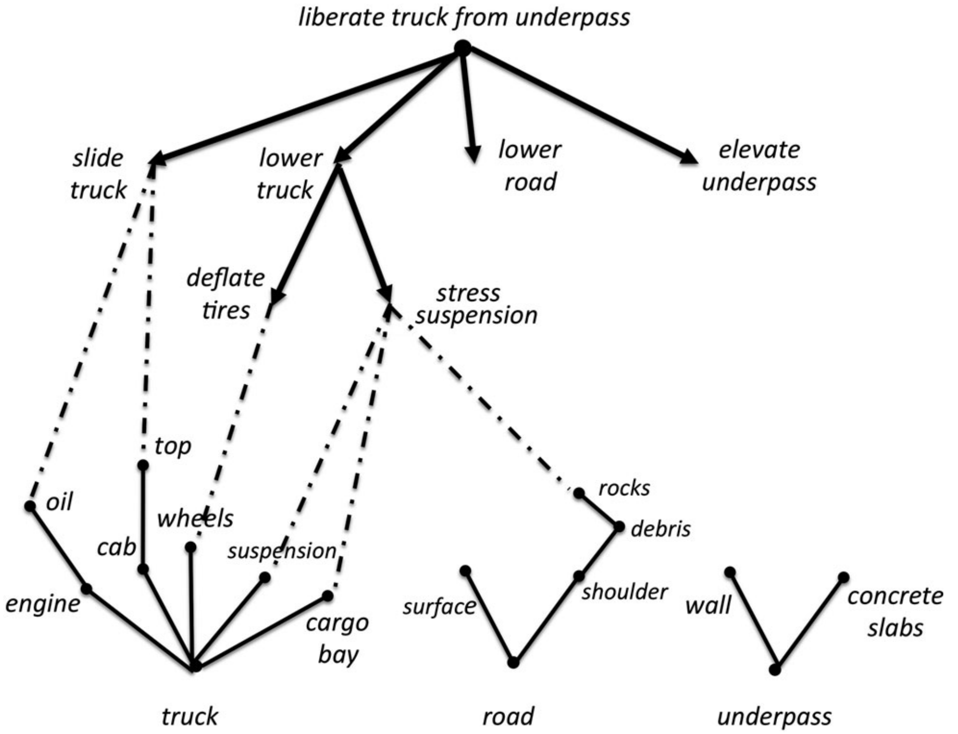

FIGURE 4. Three possible solutions for stuck truck problem.

mentioned by only one person—all other items receive a 0 (Wallach & Kogan, 1965). A variation on this method assigns points to all mentioned items that fall below a normative range of occurrences and then sums the points together (Torrance, 2008). Specifically, Milgram and Milgram (1976) assigned a 1 to every mentioned item that occurs less than 5% of the time and gives every other item a 0 before summing the scores for each item.

Objective measures of *originality* have at least three weaknesses that users should be aware of. First, *fluency* and *originality* are confounded. For example, Torrance (2008) reports a median correlation of $r = .88$ between the two. One would hope that a measure of quantity (i.e., *fluency*) and a measure of quality (i.e., *originality*) would be independent of each other, but it appears the greater the *fluency* the greater the *originality* because unique (and thus, original) responses tend to occur in longer lists of responses. Second, objective measures of *originality* have no method by which to detect responses that are irrelevant to the task and discard them from consideration. Irrelevant responses will tend to be original (i.e., other subjects do not mention them). Third, studies using larger samples tend to score original responses lower when compared to smaller samples (Silvia et al., 2008). If the originality measure is achieved by dividing by the sample size, then an original response will appear less impressive the larger the sample size (e.g., 1 response/10 people >1 response/100 people). Silvia et al. (2008, 2009) presents subjective measures of *originality* that address the three weaknesses of objective measures. In sum, BrainSwarming graphs make possible objective measures of *originality* for each graph

TABLE 1. Metrics Based on BrainSwarming Components and Creativity Measures

	Goal/sub-goals	Interactions	Features	Resources	Solution paths
Fluency (Quantity)	**Yes**, either count only leaf nodes of goal tree or all sub-goal nodes.	**Yes**	**Yes**, either count only leaf nodes of feature tree or all feature nodes.	**Yes**	**Yes**
Originality (Novelty: Statistical Infrequency)	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Flexibility (Distinct Categories Covered by Responses)	**Yes**, the category system consists of the types of solutions that are possible based on the particulars of the problem (e.g., the resources mentioned in the sub-goals).	No, but maybe someone can yet create a category system of types of interactions.	**Yes**, I have developed a category system of feature types for physical objects. Other category systems are possible.	**Yes**, resources could be categorized based on their state of matter (i.e., solid, liquid, and gas). Other categorizations are also possible.	**Yes,** one can use a category system for types of solutions based on the type of energy involved (mechanical, chemical, etc.: Hirtz et al., 2002). Other categorizations are possible.
Elaboration (Amount of Detail)	**Yes**, the intricacy of the goal network (as described in the text)	**Yes**, the intricacy of the network that leads to an interaction	**Yes**, the intricacy of the feature tree that emerges from a resource	No, resources themselves are not elaborated, but their feature trees are.	**Yes**, the intricacy of the sequence of interactions

component (i.e., *goal/sub-goals*, *interactions*, *features*, *resources*, and *solution paths*), but experimenters should be aware of the weaknesses of these measures and consider the subjective measures presented by Silvia et al. (2008, 2009).

Flexibility requires an underlying category system to classify the items listed on the graph. As category systems are constructed for the types of *goal/sub-goals*, *interactions*, *features*, *resources*, and *solution paths*, measurements of flexibility become possible. For example, Hirtz, Stone, McAdams, Szykman, and Wood (2002) created a category system of 12 energy types (e.g., mechanical, magnetic, chemical, thermal, etc.) that encompass any type of physical solution. For the *Stuck Truck* problem, all three solutions are mechanical. Placing oil on the truck roof decreases friction. Deflating tires lowers the truck. Stressing the suspension also lowers the truck. If someone created both a mechanical solution and a magnetic solution, for example, that person could be considered to exhibit more *flexibility* than a person who created only mechanical solutions.

Goals/sub-goals can be categorized in various ways: by the resources mentioned in the sub-goal (e.g., *lower truck*, *lower road*, and *elevate underpass*), whether the direction of the change is up or down (e.g., *lower* or *elevate*), etc. The categorizations possible all depend on the particulars of the problem. Currently, a category system has not been built for the types of *interactions*. Consequently, a *flexibility* measure cannot yet be applied to this element of a BrainSwarming graph. A simple category system exists for *resources* based on their state of matter (e.g., solid, liquid, or gas), but other categorizations are possible. A detailed category system for *features* has been constructed. Building on the Russian problem solving system called TRIZ (Altshuller, 1994, 1999), I have developed an extensive category system consisting of 50 types of features (McCaffrey, 2011) possessable by any physical object (e.g., color, weight, motion, size, aroma, texture, symmetry, spatial relations, etc.). Such a category system makes possible the measurement of *flexibility* of a person's responses across the many types of features that an object might possess.

Furthermore, the category system of features can be used to reveal what people tend to overlook about an object and prompt them to be more creative with the object. McCaffrey and Spector (2012) has shown that during feature listing tasks for common objects, people overlook on average 65% of the possible feature types. For each object, however, different feature types are noticed. For example, for a candle, people tend to overlook motion—no subject of mine has ever listed a feature describing how candles are motionless when they burn. However, for a basketball, motion is commonly noticed (e.g., bounce, spin, shoot, and roll). Importantly, the overlooked feature types can be leveraged to build innovative designs. For example, the *self-snuffing candle* was created by leveraging two overlooked features: weight (i.e., candles lose weight when they burn) and motion (McCaffrey & Spector, 2012). Place a candle on one side of a scale-like structure with a counterbalanced weight on the other side. The candle will slowly rise as it burns due to its weight loss. A snuffer is placed above the candle so it will slowly move into the snuffer and extinguish itself. Two candle companies verified that nine of 10 of our new designs made from overlooked features, including the *self-snuffing candle*, were indeed novel to the candle industry (McCaffrey & Spector, 2012). Furthermore, McCaffrey and Spector (2012) has shown that examining the extensive category system of features is effective for overcoming the cognitive obstacle *design fixation* (Jansson & Smith, 1991), which occurs when designers want to create novel designs but fixate on the features of known designs they have already perceived.

Elaboration can be easily quantified for goal trees, feature trees, and solution paths. In Figure 4, there are three non-elaborate solution paths shown, each involving only one interaction. Each solution path could be presented in more detail and thus made more *elaborate*. For example, Figure 4 does not show how to deflate the tires, which could require an unmentioned resource (e.g., a screwdriver or other tool) to press the air valve on the tires. The measure of *elaboration* for a solution path could be quantified by adding together the number of resources and features involved in the interactions related to achieving the sub-goals in a solution path. The more interactions, resources, and features involved, the more *elaborate* the solution is considered to be. A similar measurement process would apply to goal trees and feature trees. *Elaboration* cannot be applied directly to resources, but rather to feature trees associated with the resources.

EXAMPLE USING THE METRICS

Figure 5 shows the graphs created from the activity of two hypothetical subjects working on the *Stuck Truck* problem. Comparing their performances, both subjects crafted a single solution. Subject 1 deflated the tires to lower the truck a few inches, while Subject 2 put the truck's oil on the top to try to slide it from beneath the underpass. Although both subjects created one solution, the graphs reveal that Subject 2 explored three other possible sub-goals while Subject 1 while did not. As Table 2 shows, Subject 2 had a higher fluency score for sub-goals than Subject 1. Furthermore, Subject 2's sub-goals included all three possible resources (i.e., truck, road, and underpass) while Subject 1 only considered the truck. Thus, Subject 2 was given a higher flexibility score (three resources mentioned in sub-goals/3 resources) than Subject 1 (one resource mentioned in a sub-goal/3 resources). Subject 2 broke the resources down into many features (i.e., 12 total) while Subject 1 did not, as shown by a higher fluency score for features. For this initial example analysis, I did not calculate *originality* and *elaboration* scores.

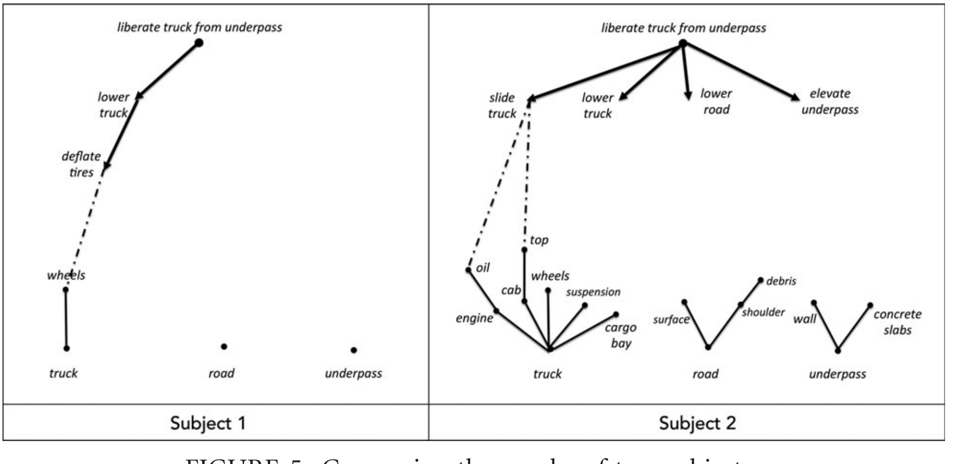

FIGURE 5. Comparing the graphs of two subjects.

TABLE 2. Comparing Two Subjects on the *Stuck Truck* Problem

	Goal/ sub-goals	Interactions	Features	Resources	Solution paths
Fluency (Quantity)	1 vs. 4	1 vs. 1	1 vs. 12	3 vs. 3	1 vs. 1
Flexibility (Distinct Categories Covered by Responses)	1/3 vs. 3/3 Use the category system of sub-goals involving the different resources: truck, road, and underpass.	NO CATEGORY SYSTEM has been developed for interactions.	1 part vs. 12 parts All features listed by both subjects could be considered parts.	0 vs. 0 Neither subject listed more resources than were initially given with the problem.	1 mechanical solution vs. 1 mechanical solution.

In sum, the proposed graphing method reveals multiple differences between the two subjects' performances beyond whether or not they each proposed a plausible solution to the problem. In this brief analysis, I used only nine of the 18 proposed metrics made possible by crossing graph elements with creativity measures. Overall, I discovered that Subject 2 explored more possible solutions and considered many more features than Subject 1. Also, Subject 2 considered a more diverse set of sub-goals. These results revealed themselves in the *fluency* and *flexibility* scores used.

NEW METRICS INVOLVING PERCENTAGES

Thus far, I have proposed many metrics based on adapting the four traditional creativity measures (i.e., *fluency*, *originality*, *flexibility*, and *elaboration*) to the five main components of a BrainSwarming graph (i.e., *goal/sub-goals*, *interactions*, *features*, *resources*, and *solution paths*). Of the 20 possible combinations, 18 of them result in solid metrics (see Table 1). Further metrics can be added by considering relationships between two existing metrics. Several candidate percentages include the following. First, the percent of solution paths compared to the number of sub-goals mentioned gives an indication of how many sub-goals actually led to a solution. If the percentage is high, then perhaps the problem solvers are too cautious in that they are only considering sub-goals in which a solution is readily apparent. If nearly every sub-goal listed resulted in a solution, then perhaps insufficient risk was taken by the problem solvers—they did not fail enough because they did not think wildly and broadly enough. Second, the percent of resources that end up in a solution path is another possible metric. If the percent is low, then perhaps the problem solvers have completely ignored or prematurely dismissed promising resources. At least two other percentage metrics are possible. Third, what percent of features became part of a solution path? Fourth, what percent of interactions became part of a solution path?

In sum, four additional metrics involving percentages brings the total number of new candidate metrics for insight problem solving to 22.

FORMING A CREATIVITY QUOTIENT FOR INSIGHT PROBLEM SOLVING

Adding together different weighted linear combinations of these 22 metrics creates a family of candidate quotients for an overall measure of insight problem solving ability. For example, if one wanted to emphasize usefulness, then perhaps weight the number of solution paths more heavily than the other quantities. To measure "raw" creativity without regard for usefulness, then perhaps ignore the number of solutions entirely (i.e., give it a weight of zero). These are just two of the ways to weight the different metrics and add them together. Future research and extensive dialog among creativity researchers will determine which quotients from this family are most informative and diagnostic.

VERBAL PROTOCOL TO GENERATE BRAINSWARMING GRAPHS

As solvers do not generally know how to create BrainSwarming graphs, experimenters need to initially create them until solvers are trained in their construction. Solvers will be asked to talk aloud while problem solving (i.e., use a verbal protocol: Ericsson & Simon, 1993) and be recorded (i.e., audio or video). An experimenter will be present in the room only to encourage solvers to continue talking when they become silent. After the experiment, the experimenter will construct a BrainSwarming graph based on the recording and use it to calculate the desired metrics.

As verbal protocol has generally been found not to hinder problem solving performance (Ericsson & Simon, 1993; see also Bloom & Keil, 2001; Chrysikou & Weisberg, 2005; Dominowski, 1998; Reisberg, 2000; Taylor & Dionne, 2000), this process of using a verbal protocol in the construction of BrainSwarming graphs should not interfere with a solver's performance.

DIAGNOSES AND PRESCRIPTIONS

Once the desired metrics from a solver's BrainSwarming graph (or graphs) have been calculated (or perhaps averaged across several graphs), one can then engage in the process of diagnosing cognitive inhibitors to problem solving performance and prescribing targeted counter-measures. In this section, I present several known cognitive obstacles—including recent obstacles that BrainSwarming graphs helped to uncover. I also present their effective counter-measures and how the counter-measures are visually presented in a BrainSwarming graph.

McCaffrey (2012) presented a new theory of innovative problem solving, the obscure features hypothesis (OFH), which articulates how any innovative solution is based on noticing and then using a feature of the problem that is commonly overlooked (i.e., obscure). Unlike previous approaches, the OFH makes possible a systematic way to improve innovation in problem solving (McCaffrey, 2012). In contrast, the *representation change view* (Knoblich, Ohlsson, Raney, Haider, & Rhenius, 1999; Ohlsson, 1992) states that solvers initially form an incorrect or incomplete representation of the problem at hand. As no category system of possible representation changes has yet been articulated, this approach has not led to a systematic way to help solvers execute various types of representation changes needed to improve their performance. The *distant association view*

(Mednick, 1962) states that solutions to insight problems are based on concepts that are distantly associated to the concepts of the problem. However, no category system of types of associations has yet been developed and no method of proceeding to the associations at the proper distance has been devised. Thus, this approach has not led to a systematic method of exploring the space of associations at the proper distance.

The systematic approach made possible by the OFH states that each cognitive obstacle should be characterized by the types of features that it hides from attention and an effective countering technique should be crafted to uncover the obscure members of those concealed types of features. McCaffrey (2012) presented the first highly effective technique for countering the classic obstacle to innovation: *functional fixedness* (Duncker, 1945). *Functional fixedness* is traditionally characterized as fixating on the typical use of an object or one of its parts (Duncker, 1945). Following the OFH approach, McCaffrey (2012) redefined *functional fixedness* as the tendency to ignore four types of features of objects: parts, material, shape, and size. The generic-parts technique (GPT) uses a graphical method to systematically redescribe an object more generically in terms of its underlying parts, material, shape, and size. People trained in GPT solved 67.4% more insight problems than a control group (McCaffrey, 2012).

Since the 2012 study, another known cognitive obstacle (i.e., *design fixation*: Jansson & Smith, 1991) has been redefined in terms of the feature types that are overlooked; and an effective countering technique (i.e., the *feature-type spectrum*) has been proposed and successfully tested (McCaffrey & Spector, 2012). Furthermore, the OFH approach has helped to identify new cognitive obstacles to innovation (i.e., *narrow verb associations*, *assumption blindness*, and *analogy blindness*) and their effective countering techniques (McCaffrey & Krishnamurty, 2014; McCaffrey & Spector, 2011a, 2012; McCaffrey et al., 2014).

BrainSwarming graphs have been instrumental in terms of visualizing and then identifying new cognitive obstacles. The BrainSwarming graph (Figure 1) portrays four basic zones where things can be overlooked. Moving from top to bottom of the graph, things about the *goal* can be overlooked; as well as things about *interactions*, *features*, and *resources*. *Functional fixedness* (Duncker, 1945) is one mental obstacle that hides some features of the resources; *design fixation* (Jansson & Smith, 1991) is another. *Narrow verb associations*, *assumption blindness*, and *analogy blindness* all shield aspects of the goal from one's attention.

Figure 6 shows the kind of diagram used for the GPT (McCaffrey, 2012), which successfully counters *functional fixedness* (Duncker, 1945). Figure 7 shows the execution of the GPT in a full BrainSwarming graph. Figure 7 also shows the obstacle *narrow verb associations* and the execution of its counter-technique (McCaffrey & Spector, 2012).

For example, consider the following insight problem: the *two rings problem* (McCaffrey, 2012). Two heavy steel rings need to be securely fastened together in a figure-eight configuration using only a long candle, a match, and a 2-inch cube of steel. A wax bond will not fasten the rings securely enough so the solution depends on noticing that the wick is a string that if extricated can tie the two rings together.

The GPT systematically deconstructs an object so that new uses can emerge. Two questions are repeatedly asked as a parts tree (Figure 6) is constructed. First, "Can I decompose this further?" If so, the solver should break it down further into smaller parts and build another level of the parts tree. Second, "Does the current description imply a

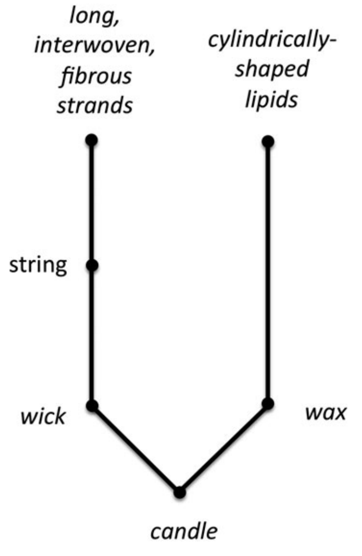

FIGURE 6. Generic parts tree for a candle.

use?" If so, the solver should add another level of the tree by rewording the description more generically using its material and shape. Together, the GPT focuses the solver on parts, material, shape, and size (the parts get smaller as the tree is built). Figure 6 shows a parts tree for a candle.

Working bottom-up on a full BrainSwarming graph, Figure 7 shows the GPT applied to the candle. Since the *two rings problem* becomes trivial once *string* is uncovered (McCaffrey, 2011), I only created the parts tree to that point. Furthermore, to avoid clutter in Figure 7, I did not show parts trees for the other resources.

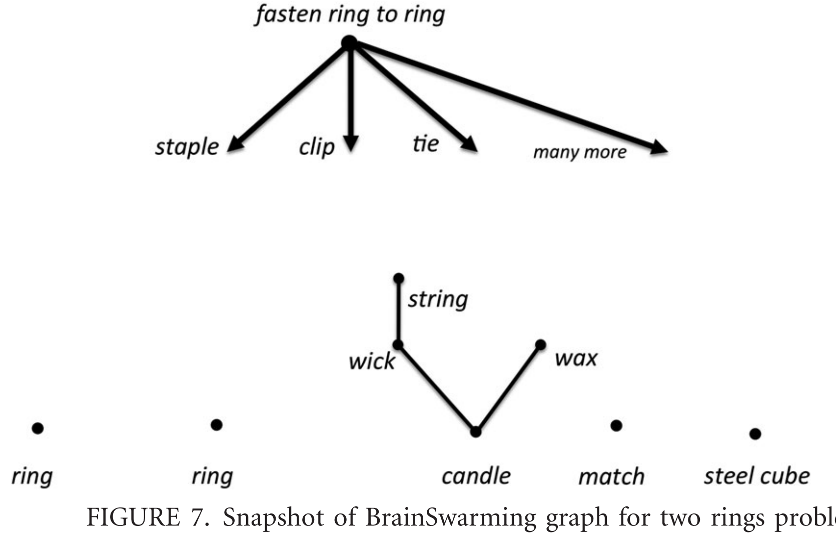

FIGURE 7. Snapshot of BrainSwarming graph for two rings problem.

Working top-down in Figure 7, any goal can be phrased so that it starts with a verb (Hirtz et al., 2002). The verb expresses either the change that needs to be enacted to accomplish the goal or points to the fact that change is to be avoided (e.g., *maintain*, *preserve*, and *sustain)*. In Figure 7, *fasten* describes the desired change. More specific synonyms of *fasten* (i.e., hyponyms) imply concrete ways to *fasten* things together (e.g., *staple*, *clip*, *tie*, etc.). An online thesaurus, WordNet (Miller, 1995), lists 61 different hyponyms for *fasten*. McCaffrey and Spector (2012) showed that WordNet contains significantly more hyponyms and synonyms than subjects can list on their own—a human limitation called *narrow verb associations*. For six common verbs, WordNet contains many hyponyms: *fasten* (61), *remove* (172), *guide* (50), *transport* (46), *mix* (24), and *separate* (115). For these verbs, however, subjects were able to list only 8.1 synonyms on average with a margin of error of 2.8. Furthermore, of the 8.1 synonyms, 3.9 were hyponyms. Quite possibly, WordNet could help problem solvers find more concrete ways to achieve a goal than they can devise on their own.

For the *two rings problem*, the verb *tie* is one of the hyponyms of *fasten*, so using the WordNet thesaurus to flesh out the goal network could provide an important clue to problem solvers.

Figure 8 shows the sequence of interactions that will solve the *two rings problem*. Once you notice that the wick is a string, you then need to free the string from the wax. Figure 8 indicates that this sub-goal was achieved by interacting the wax with the cube of

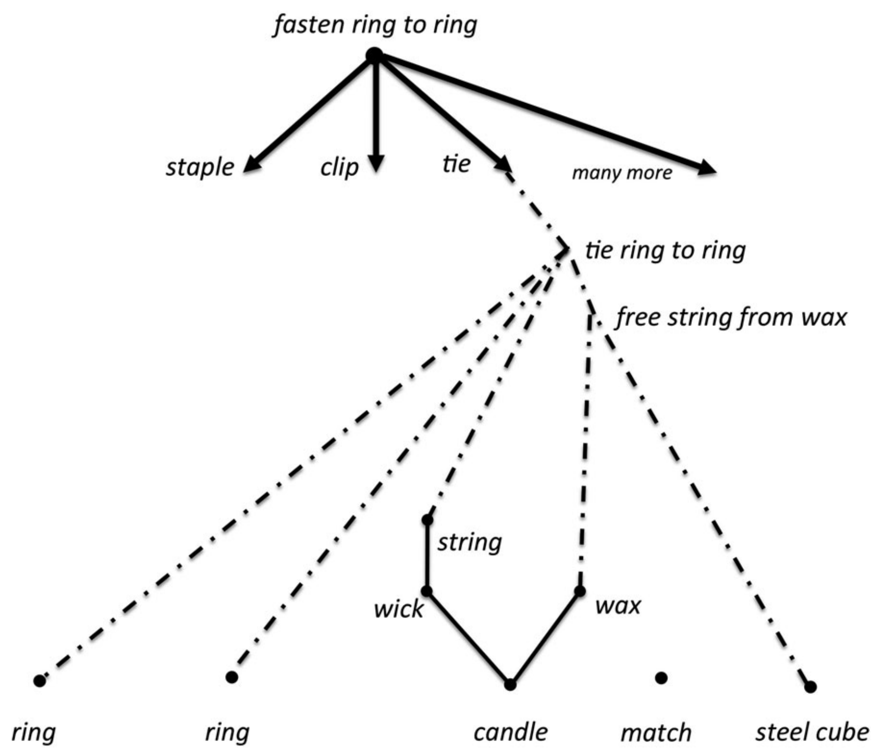

FIGURE 8. Final BrainSwarming graph for two rings problem.

steel. This solution path could be made more elaborate by being explicit about scraping the wax against the edge of the steel cube. Other ways, such as scraping the wax against a table's edge, could also be employed to accomplish this sub-goal.

Once the string is free from the wax, then it can be used to tie the rings together, which, in effect, fastens the rings together. Artificial intelligence software exists that can solve this problem once the program learns that wicks are made of string (McCaffrey & Spector, 2011a,b). The software can find sequences of interactions to satisfy the goal once a human operator helps it uncover the crucial obscure features of various resources (McCaffrey & Spector, 2011a,b).

In sum, the BrainSwarming graph visualizes the major zones of any problem (i.e., *goal/sub-goals*, *interactions*, *features*, and *resources*). Classic inhibitors (e.g., *functional fixedness*) and their counter-measures (e.g., GPT) each have their place in the graph. This approach has helped uncover new inhibitors (i.e., *narrow verb associations*, *assumption blindness*, and *analogy blindness*) and their counter-measures (McCaffrey & Krishnamurty, 2014; McCaffrey & Spector, 2012; McCaffrey et al., 2014). Examining the metrics that emerge from a BrainSwarming graph could lead to a detailed diagnosis of an individual's (or team's) weaknesses that could then be corrected by prescribing the proper counter-measures.

CONCLUSIONS

A BrainSwarming graph visualizes problem solving activity for insight problems. This graphical method permits the creation of at least 22 new quantitative metrics. From these metrics, researchers can diagnose the creative weaknesses of individuals and teams. Multiple creative weaknesses are known and, for most known weaknesses, there exists an effective counter-measure. This proposed measurement system for insight problems is, at this stage, theoretical and will require the empirical testing by many creativity researchers to determine its overall usefulness. Future research will help determine which of the 22 problem solving metrics are most informative and diagnostic. Future research is also needed to form a useful composite quotient of overall insight problem solving ability, one most likely based on a linear combination of some subset of the 22 metrics. Creativity researchers will undoubtedly identify new creative weaknesses not addressed in this paper and suggest alterations to the present framework. In sum, in my estimation, the BrainSwarming graph structure will make possible an important leap in our ability to systematically quantitate, diagnose, and improve performance on insight problems.

Tony McCaffrey is the sole author of this article and is responsible for its content.

REFERENCES

Altshuller, G. (1994). *And suddenly the inventor appeared*. Translated by Lev Shulyak. Worcester, MA: Technical Innovation Center.

Altshuller, G. (1999). *The Innovation Algorithm: TRIZ, systematic innovation, and technical creativity*. Worcester, MA: Technical Innovation Center.

Bloom, P., & Keil, F.C. (2001). Thinking through language. *Mind & Language*, *16*, 351–367.

Chrysikou, E., & Weisberg, R. (2005). Following the wrong foot- steps: Fixation effects of pictorial examples in a design problem- solving task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1134–1148.

Dominowski, R.L. (1998). Verbalization in problem solving. In D.J. Hacker, J. Dunlosky & A.C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 25–45). Mahwah, NJ: Erlbaum.

Duncker, K. (1945). On problem-solving. *Psychological Monographs*, *58*, 1–113.

Ericsson, K., & Simon, H.A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

Gallupe, R.B., & Cooper, W.H. (1993). Brainstorming electronically. *Sloan Management Review*, *35*, 27–36.

Gallupe, R.B., Dennis, A.R., Cooper, W.H., Valacich, J.S., Bastianutti, L.M., & Nunamaker, J.F. (1992). Electronic brainstorming and group size. *Academy of Management Journal*, *35*, 350–369.

Greene, K., Thomsen, D., & Michelucci, P. (2011). *Explorations in massively collaborative problem solving* (pp. 716–719). Proceedings of the IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing.

Guilford, J.P. (1967). Creativity: Yesterday, today, and tomorrow. *Journal of Creative Behavior*, *1*, 3–14.

Harvard Business Review (Producer) & McCaffrey, T. (Writer) (2014). BrainSwarming: Because brainstorming doesn't work [Video]. http://hbr.org/video/3373616535001/brainswarming-because-brainstorming-doesnt-work

Hirtz, J., Stone, R., McAdams, D., Szykman, S., & Wood, K. (2002). A functional basis for engineering design: Reconciling and evolving previous efforts. *Research in Engineering Design*, *13*, 65–82.

Jansson, D.G., & Smith, S.M. (1991). Design fixation. *Design Studies*, *12*, 3–11.

Knoblich, G., Ohlsson, S., Raney, G.E., Haider, H., & Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1534–1555.

McCaffrey, T. (2011). *The obscure features hypothesis for innovation: One key to improving human innovation*. Unpublished doctoral dissertation, University of Massachusetts, Amherst, MA.

McCaffrey, T. (2012). Innovation relies on the obscure: A key to overcoming the classic *functional fixedness* problem. *Psychological Science*, *23*, 215–218.

McCaffrey, T., & Krishnamurty, S. (2014). The *Obscure Features Hypothesis* in design innovation. *International Journal of Design Creativity and Innovation*, *3*, 1–28.

McCaffrey, T., Krishnamurty, S., & Lin, X. (2014). Cahoots: A software platform for enhancing innovation and facilitating situation transfer. *Research and Practice in Technology Enhanced Learning*, *9*, 145–163.

McCaffrey, T., & Spector, L. (2011a). Innovation is built upon the obscure: Innovation-enhancing software for uncovering the obscure. In A.K. Goel, F. Harrell, B. Magerko, Y. Nagai & J. Prophet (Eds.), *Proceedings of the 8th ACM Conference on creativity and cognition* (pp. 371–372). New York: Association for Computing Machinery. doi:10.1145/2069618.2069701

McCaffrey, T., & Spector, L. (2011b). How the Obscure Features Hypothesis leads to innovation assistant software. *Proceedings of the Second International Conference on Computational Creativity*, *1*, 120–122.

McCaffrey, T., & Spector, L. (2012). Behind every innovative solution lies an obscure feature. *Knowledge Management and E-Learning*. Special issue on Supporting, Managing, & Sustaining Creativity and Cognition through Technology, *4*, 146–156.

Mednick, S. (1962). The associative basis of the creative process. *Psychological Review*, *69*, 220–232.

Milgram, R.M., & Milgram, N.A. (1976). Creative thinking and creative performance in Israeli students. *Journal of Educational Psychology*, *68*, 255–259.

Miller, G.A. (1995). WordNet: A lexical database for English. *Communications of the ACM*, *38*, 39–41.

Mongeau, P.A., & Morr, M.C. (1999). Reconsidering brainstorming. *Group Facilitation*, *1*, 14–21.

Nielsen, M. (2012). *Reinventing discovery: The new era of networked science*. Princeton, NJ: Princeton University Press.

Ohlsson, S. (1992). Information processing explanations of insight and related phenomena. In M.T. Keane & K.J. Gilhooly (Eds.), *Advances in the psychology of thinking* (pp. 1–44). London: Harvester Wheatsheaf.

Osborn, A. (1953). *Applied imagination: Principles and procedures of creative problem solving*. New York: Scribner.

Reisberg, D. (2000). The detachment gain: The advantage of thinking out loud. In B. Landau, J. Sabini, J. Jonides & E. Newport (Eds.), *Perception, cognition, and language: Essays in honor of Henry and Lila Gleitman* (pp. 139–156). Cambridge, MA: MIT Press.

Rittel, H., & Webber, M. (1984). Planning problems are wicked Problems. In N. Cross (Ed.), *Developments in design methodology* (pp. 135–144). New York: John Wiley & Sons.

Runco, M.A., & Pritzker, S.R. (Eds.) (1999). *Encyclopedia of creativity*. San Diego, CA: Academic Press.

Silvia, P.J., Martin, C., & Nusbaum, E.C. (2009). A snapshot of creativity: Evaluating a quick and simple method for assessing divergent thinking. *Thinking Skills and Creativity*, *4*, 79–85.

Silvia, P.J., Winterstein, B.P., Willse, J.T., Barona, C.M., Cram, J.T., Hess, K.I., … Richard, C.A. (2008). Assessing creativity with divergent thinking tasks: Exploring the reliability and validity of new subjective scoring methods. *Psychology of Aesthetics, Creativity, and the Arts*, *2*, 68–85.

Simon, H. (1995). Problem forming, problem finding and problem solving in design. In A. Collen & W. Gasparski (Eds.), *Design & systems* (pp. 245–257). New Brunswick: Transaction.

Taylor, K.L., & Dionne, J.P. (2000). Accessing problem-solving strategy knowledge: The complementary use of concurrent verbal protocols and retrospective debriefing. *Journal of Educational Psychology*, *92*, 413–425.

Torrance, E.P. (1974). *Torrance tests of creative thinking: Norms-technical manual*. Lexington, MA: Ginn and Company.

Torrance, E.P. (1990a). *Torrance tests of creative thinking, norms-technical manual (figural)*. Bensenville, IL: Scholastic Testing Service.

Torrance, E.P. (1990b). *Torrance tests of creative thinking, norms-technical manual (verbal)*. Bensenville, IL: Scholastic Testing Service.

Torrance, E.P. (2008). *Torrance Tests of Creative Thinking: Norms-technical manual, verbal forms A and B*. Bensenville, IL: Scholastic Testing Service.

Wallach, M.A., & Kogan, N. (1965). *Modes of thinking in young children: A study of the creativity intelligence distinction*. New York: Holt, Rinehart & Winston.

Tony McCaffrey, Innovation Accelerator, Inc.

Correspondence concerning this article should be addressed to Tony McCaffrey, Ph.D., Innovation Accelerator, Inc., 55 Snow Road, West Brookfield, MA 01585. E-mail: tony@innovationaccelerator.com

ACKNOWLEDGMENT

This material was based upon work supported by the National Science Foundation (NSF) of the USA under Grant Nos. 1127609, 1261052, 1345439, and 1534740. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the NSF.