

Creativity Support Tool Evaluation Methods and Metrics

Tom Hewett, Drexel University
Mary Czerwinski, Microsoft
Michael Terry, Georgia Tech
Jay Nunamaker, University of Arizona
Linda Candy, University of Technology, Sydney
Bill Kules, University of Maryland
Elisabeth Sylvan, MIT

An overview and meta-analysis of psychological research on creativity

One goal of this portion of the report is to provide a brief overview of our current understanding of what the psychological research community examining creativity tells us about the topic, as well as to review some of the conceptual and methodological issues involved in the psychological study of creativity. A third goal is to discuss some of the implications of this research for requirements analysis for creativity support tools, for the design of creativity support tools, and for the evaluation of the impact of those tools intended to support creativity. This discussion is based upon a presentation and subsequent discussion at the NSF Sponsored Creativity Support Tools (CST) Workshop held in Washington, DC in June of 2005.

In providing an overview of the Psychological research on creativity we have relied heavily on various sources in *The Handbook of Creativity* (Sternberg, 1999), in particular the overview article by Mayer (1999). The authors in Sternberg's collection of reviews provide a high level view of the state of the art and findings of psychological research on creativity. The work in this Handbook is highly consistent with the work of several other authors who have also surveyed major aspects of the research findings (e.g., Csikszentmihalyi, 1997; Gardner, 1989).

Several of the authors in Sternberg adopt a working definition of creativity that is consistent with those offered by other authors and that involves several key components. Basically, creativity can be considered to be the development of a novel product that has some value to the individual and to a social group. However, it seems that the research conducted by psychologists on creativity does not allow us to clarify or simplify this definition any further. Different authors may provide a slightly different emphasis in their definition but most (if not all) include such notions as novelty and value. For example, Gardner (1989) emphasizes that creativity is a human capacity but includes novelty and social value in his definition. An important thinker and researcher on creativity, Csikszentmihalyi (1997), emphasizes that creativity involves process but stipulates that that process can be observed only where individuals, knowledge domains, and fields or social groups intersect.

In summarizing the research findings reported in the various chapters in Sternberg(1999) it is clear that that there are several diversities (Mayer, 1999) that can be thought of as being underlying dimensions to creativity and the study of creativity. For example, one dimension is that creativity can be a property of people, a property of products, and a property of a set of cognitive processes. This diversity leads to a concern with individual differences between people. It also leads to a concern with the properties of a product that make it novel and valuable. Finally, it leads to a concern with analyzing the steps and processes of thinking which are associated with production of a creative result.

A second dimension of creativity and in creativity research to be found in various chapters in Sternberg (1999) is that creativity can be thought of as a personal and a social, societal or cultural phenomenon. At an individual level creativity is said to involve the development of something novel and valuable to the individual. At the social level it involves a creation that adds something new to the culture. This dimension parallels the distinction made by Boden (1990) between P-creative and H-creative. Boden's important conceptual clarification helps advance the discussion of creativity as it then becomes clear that an individual may be personally creative in coming up with something novel to themselves (P-creative), without necessarily being H-creative by making a contribution to the human race.

A third dimension in creativity research and to creativity to be found in Sternberg (1999) is that creativity can be thought of as being common or frequent, or it can be thought of as being rare. Effectively, some aspects of the research on creativity suggest all humans are potentially capable of creativity (in Boden's P-creative sense). Alternatively the research suggests that major creative works (in Boden's H-creative sense) are rare.

Another dimension of creativity to be found in the literature on creativity research discussed in Sternberg (1999) is that creativity may involve domain specific characteristics but that there are also domain independent or general phenomena as well. In other words, there appear to be general skills associated with being creative that apply across a variety of situations or domains of knowledge and/or practice. On the other hand it seems that different domains require extensive domain knowledge and domain specific special abilities (e.g., the physical skills required by sculpting are different than those required by composition of music).

The fifth dimension of creativity research and creativity to be found in Sternberg (1999) is that creativity can be seen as being quantitative or it may be seen as being qualitative. For example, individuals may have varying amounts of creativity (e.g., as measured by psychometric tests). Furthermore different people may display different types of creativity (cf., Gardner, 1986).

Yet another dimension of creativity and in creativity research to be found in Sternberg (1999) is that creativity can be individual or it can be social, in the sense of a group of people working together. For example, it is possible to study how individuals may be creative or produce a creative result. Similarly, groups of people working together may also produce a creative result that is a group result and that is not uniquely the product of a single member of the group. Thus it becomes necessary to study social entities, social products and social processes to fully understand creativity.

Recognition of the fact that there are multiple dimensions to creativity and in creativity research, leads us to propose that these various aspects of creativity research and creativity should be thought of as being different dimensions of a taxonomy for creativity studies and creativity support tools. In other words, the problem of developing Creativity Support Tools is one in which one must first decide in which intersection of the n-dimensional taxonomy one wishes to study and work.

For example, a Creativity Support Tool might be designed to support group work either by focusing on facilitating processes through to enhance creativity or by enabling the production of a physical artifact that is both novel and useful. This tool might not be of any use at all to an individual. While it is clear that not all possible permutations of this n-dimensional taxonomy have been explored, it does seem safe to argue that it should be possible clarify future discussions of Creativity Support Tools if investigators make use of such a Taxonomy to clearly stipulate which particular intersection of factors best characterizes the goals and nature of the Creativity Support Tool upon which they are working. This specification would also be of assistance in deciding upon which methodologies and metrics should be used in assessing the degree to which a Creativity Support Tool is thought to facilitate creative work.

An over view and meta-analysis of research methods used to study creativity

Mayer (1999) provides a review and some observations on a variety of behavioral science research methods used to study creativity. Specifically he discusses psychometric methods, experimental methods, biographical methods, biological methods, computational methods and contextual methods. In addition he addresses some of the strengths and weaknesses of these various methods, making the important point that in studying creativity one should use more than one methodology to allow one to compensate for the weakness of a single methodology. This can be done by bringing to bear the strength of another methodology and is based upon the assumption that the two different sets of weaknesses can effectively cancel each other out. The goal of this section of the report is to summarize, and, in a few cases, comment on Mayer's observations. It should be also noted that this idea of use of multiple converging methods and metrics for studying creativity and creativity support tools motivated much of a breakout group discussion described below.

The first set of research methods addressed by Mayer is the Psychometric method. This collection of procedures involves the development of various psychological tests that are intended to assess various traits or characteristics of creative people. The strength of this method, Mayer argues, is the fact that it has a long history and the procedures for developing such tests are well established. The major weakness of the Psychometric method, however, is basically that such tests, e.g., tests of divergent thinking, don't seem to predict creative thinking. Another way of thinking about this weakness is that Psychometric tests lack predictive validity, criterion validity and discriminant validity. That is, they don't seem to offer a strong way of allowing us to identify who will be creative, what results will be accepted as creative, or allow us to distinguish between creative and non-creative people.

The fundamental nature of the problem here can be understood by noting that even if one has a significant correlation between the pencil and paper test and some other indicator of creativity, the real power of a relationship between two variables lies not in a statistically significant correlation but in the percentage of variance accounted for by the relationship. If one takes the square of the correlation coefficient one has then calculated the percentage of variance accounted for by knowing the score on one variable when trying to predict a value on the other. Thus a significant correlation of 0.5 only accounts for 0.25% of the variability in the score on the other variable.

Mayer (1999) describes experimental methodologies as having the inherent strength of all laboratory research. The control one exercises increases the validity or strength of one's conclusions. (Effectively, in a well controlled experiment the researcher has only two possible explanations for the result of a manipulation. Either the result is an effect of the manipulation or it has occurred by chance. Careful selection of a level of statistical significance allows one to estimate the possibility of a chance result.) Not surprisingly, the weakness of experimental methods identified by Mayer is that with increased control comes reduced generalizability of one's conclusions. This is a long standing and well understood problem with experimental research which has been grappled with in a variety of ways (e.g., Campbell & Stanley, 1966; Cook & Campbell, 1979; Webb et al, 1969; Webb et al, 1981). One important implication of this inescapable trade-off between control and generalizability is that laboratory definitions of "creativity" are often so tightly constrained that they do not capture more than a piece of a person, product or process. And that piece is usually observed outside of a natural context.

Turning to biographical methodologies, Mayer notes that their strengths derive from the fact that carefully documented histories can provide both detail and a feeling of authenticity. However, not all histories are carefully documented at the time events are taking place. Furthermore, there are potential biases introduced as a result of a focus on a small set of pre-selected people. Both of these problems raise the concern that one may have access to only part of the data to work with or that the reporting of events may

be influenced to some degree by selective memory. For example, autobiographical accounts composed years after the fact may talk only about successes and may omit or downplay failures, etc.

In recent years a whole new collection of biological methodologies have become available, and as Mayer points out, brain event recording data provides information not available to other methodologies. The down side he notes is that it is not clear yet how cognitive activity can be reduced to brain activity or vice versa. We interpret this to mean that reports of brain behavior correlates provide us with some new information but as yet it is not clear that this will lead to any new understanding of the phenomena of creativity.

Another relatively recent set of methodologies for the study of creativity involves computational modeling. In computational modeling one instantiates a theory in the form of a computer executable program that is thought to incorporate the same types of constraints and limitations as found in human cognition. Mayer points out that these modeling efforts allow for a rare level of precision which allows for objective testing via simulation. The weaknesses he identifies include the observation that it is not clear that such modeling will ever have a broad enough scope to deal with the full range of phenomena of interest.

In addition Mayer (1999) claims that computational modeling assumes that cognition can be reduced to mathematics. On this point we tend to disagree with Mayer for several reasons. First it is quite possible to use a digital computer to simulate certain analog processes and produce output that simulates and is indistinguishable from an analog device. Second, it is quite possible to use a digital computer with a single CPU that processes events in serial order to simulate certain parallel processes and produce output that is indistinguishable from a parallel device. Finally, there appears to be no a priori reason why a digital computer can not simulate cognitive events and processes in such a way that it produces output which is indistinguishable from the human. The fact that a mathematically based device is used to simulate another device does not entail the second device is thought to be reducible to mathematics, only that mathematics can be used to model it.

The final set of methodologies for the study of creativity discussed by Mayer (1999) are contextual methodologies. The strength of these methodologies lies in the fact that they place the study of creativity in a personal, social, societal, cultural and even an evolutionary context. The projects studied are defined by the practitioner and the research studies creativity using research based in actual practice. The weakness of these methodologies identified by Mayer is the shortage of data and of testable theories based on such studies. Another weakness of these methodologies is related to the sources of strength and weakness discussed in experimental methodologies. The further one moves away from the controlled laboratory situation the more difficult it becomes to establish a clear unambiguous set of relationships that support valid conclusions. Research based in actual practice often supports many alternative explanations of what happens and how it happens.

Is creativity enhancement actually a reasonable goal?

All of the complexities one encounters in trying to study creativity, e.g., deciding upon which part of the n-dimensional space one wishes to explore and then deciding upon the appropriate methodologies and metrics to employ, reveal the rich set of design issues that developers of creativity support tools face. The psychological literature provides no clear, unequivocal answer to whether or not creativity can be enhanced. There are many different variables that have been proposed as having a role, including individual abilities, interests, attitudes, motivation, intelligence, knowledge, skills, beliefs, values and cognitive styles. Thus it seems that individual, social, societal and cultural differences and factors may all matter, at some time or another and under some circumstance or another.

Despite all these complexities, there is some hope. As Hewett (in press) has argued, we may not yet understand enough about the conditions under which creativity is going to happen to be able to help it along, but we do understand some of the things that make it harder for creativity to happen. Knowing what disrupts creativity makes it possible to figure out ways of staying out of the way and not interfering. Furthermore, Nickerson (1999) and Csikszentmihalyi (1997) have provided useful evidence suggesting that there are techniques useful in teaching and/or enhancing personal creativity. Both of these authors show strong convergence in their conclusions. Furthermore many of the techniques identified by Nickerson that may be useful in teaching creativity can also be applied as personal improvement techniques. A list of factors involved include such things as: Establish the purpose and intention of being creative; Build basic skills, Encourage acquisition of domain-specific knowledge, Stimulate and reward curiosity and exploration; Build motivation; Encourage confidence and risk-taking; Focus on mastery and self-competition; Provide opportunities for choice and discovery; and Develop self-management (meta-cognitive) skills.

Does this list of strategies for teaching or improving personal creativity have implications for design and development of Creativity Support Tools? The answer appears to be “yes.” As pointed out by Linda Candy (personal communication) there are three clear HCI objectives implied by what we do know. The first objective is to enhance the personal experience of the person who wants to be creative. The second is to look for ways to improve the outcomes and artifacts. The third objective is to support the improvement of process by providing tools that are designed with certain functional requirements in mind.

The types of functional requirements and or design criteria which should be useful have been articulated in a series of papers by Candy and Edmonds (1994, 1995, 1996, 1997) and further extended and explored by Hewett (in press). For example, any Creativity Support Tool should allow the user: to take an holistic view of the source data or raw material with which they work; to suspend judgment on any matter at any time and be able to return to that suspended state easily; to be able to make unplanned deviations; return to old ideas and goals, formulate, as well as solve, problems; and to re-formulate the problem space as their understanding of the domain or state of the problem changes.

Breakout group report on Evaluation of CST

After a series of presentations earlier in the workshop a group of people concerned particularly with the problems of appropriate methodologies and metrics for assessing the impact of computer-based tools intended for creativity support met to discuss some of the issues raised by the workshop presenters. Not surprisingly, with very little discussion the group found they shared an overriding belief in a fundamental principle of evaluation of support tools. Basically this principle states that there is a FAMILY of evaluation techniques which must be brought to bear in order to CONVERGE on the key issues involved in the study of Creativity Support Tools.

The reasons for this principle lie in the fact that any research methodology will inherently have some fundamental assumptions and/or flaws which may make it inappropriate as a single tool for a thorough, meaningful interpretation of a result. For example, evaluation techniques range from lab-based, controlled studies (perhaps only possible when you know an area well or it is mature, or when the area is extremely new) to field studies (performed early on to understand the problems and the corresponding scope) to surveys (which have only limited value) and deep ethnographies (which are powerful methods for understanding user behavior and generating hypotheses).

The group also agreed upon an important corollary of the principle that multiple methods are required for

assessing different aspects or different people, processes, or products involved in creative work, i.e., converging lines of evidence are needed. This corollary is that no single metric or measurement technique is without assumptions and inherent error of measurement. Thus various measurement techniques must, wherever possible, be combined with other metrics or measurement techniques in ways that the different measure converge upon an answer to the question(s) being asked in evaluation. The additional complexity raised by consideration of the need for multiple evaluation methods and multiple evaluation measures was that no single method or measure will be appropriate for all situations or all aspects of the complex phenomenon of creativity.

That being said, the breakout group turned its attention to generating examples of the types of questions that should be asked in evaluation studies of Creativity Support Tools (recognizing that the list is not exhaustive and that not all questions would be addressed in a single study). The list of questions generated by the group includes:

Example/sample questions to be asked in evaluation:

- Is this technique better than existing practice?
- Does it expand its use to other contexts?
- Have you learned how to improve this tool based on this evaluation?
- How does the tool/technique influence the creative process?
- What facets of creativity are affected? To what degree?
- How brittle is the tool/technique? How accepted is it by the users over the long term?
- Does it celebrate diversity?
- How does this method complement others in the family of tools/techniques?
- What is the task-to-technology “fit”?

The next aspect of the breakout group’s discussion involved producing examples of possible metrics and measures that might be used to answer one or more of these questions. Recognizing that this list is neither exhaustive and that not all of these metrics and measures would be appropriate in all evaluation studies, the list generated by the breakout group consisted of the following:

Sample measures and metrics to be used in an evaluation:

- #of unique alternatives attempted
- Degree of radicalism/conservatism if alternatives attempted
- Value of solutions attempted (to whom?)
- Quality of solutions (Gary Olsen has developed expert-derived scale based on what it is you are creating)
- Side Effects: serendipitous solutions
- Time to come up with solutions
- Satisfaction with solutions
- Progress if there is a deadline imposed
- Tradeoffs or “cost-benefit” analyses of cognitive resources applied or allocated to solutions v. those freed by alternative techniques (the last two or three can be applied “socially” as well as individually)
- Organizational agility
- # of Person Hours required for satisfactory solution
- # of people supported by the tool/process
- Cultural appropriateness of the tool/process
- People’s subsequent buy-in of the tool/process
- Ease of learning and remembering

of errors made through the user interface while using the tool/process

In the final stages of discussion the breakout group turned to looking at some of the case study examples presented during the workshop or discussed during this breakout session (with thanks to Michael Terry, Gary Olson and Jay Nunnamaker). Beyond the superficial generalization that creativity support is difficult, and evaluation of its success is complex, it was possible to summarize some important general lessons from these presentations.

Commentary on the Breakout group report.

Reflection on the breakout group discussion and the case study examples discussed there basically provides a set of guidelines for planning and conducting the process of developing and evaluating a Creativity Support Tool.

Step 1: First it is necessary to observe the activities and problems users are having in real time, either through field/ethnographic research, computer logging, or via actual participatory design.

Step 2: Next the researcher must gather user requirements for design of a system that solves real user problems or assists them in activities that they need to perform.

Step 3: Design and implement a solution. (This can often be done quickly and inexpensively using low fidelity prototypes.)

Step 4: Iterate via a series of evaluation studies that might start out qualitatively but end up being quantitatively examining new tools versus existing practice.

Step 5: Repeat Steps 3 and 4 as often as needed.

Step 6: Finally, follow up with the end system out in the field longitudinally, with users using the tools to do their real work in real time over an extended period of time.

Table 1 summarizes the techniques, what they are good for, and advantages and disadvantages.

Technique	Good For	Advantages	Disadvantages
Controlled study	For specific questions and when you know an area is ripe for improvements.	Rigorous.	Time consuming. Low external validity.
Field Study	Early on to understand the problems and the corresponding scope		
Survey	A quick overview or description of a phenomenon.	Quick, easy to administer and analyze.	Limited value. Self-report.
Deep ethnography	when you have no idea about the real problems to solve or what a very deep understanding of an area		Most time-consuming

Table 1. Various research methods for studying creativity and their pros and cons.

For some additional sources of information on several of the ideas expressed here see the case study descriptions provided below and the papers by Campbell & Fiske (1959) and Hewett (1986).

When the breakout group report was circulated for comment there were some very useful elaborations suggested. For instance, it was thought that Step 1 above is particularly important and challenging -- and in the context of creativity research may be more so than in other domains. Many creativity tasks are loosely defined at best. The strategies and tactics that people who are working creatively bring to a task are situation-specific and often ideosyncratic. Users may apply tacit knowledge and have difficulty articulating the task without performing it. This makes step 1 especially important. In other domains, where activities are more formalized, you can often collect valid requirements even if you skip or minimize step 1, but that is likely to be riskier for creativity support tools. For similar reasons, during step 1 it is especially important to "tease out" higher-level thinking that occurs. Log studies can contribute, but they can't replace research that listens to users, asks "why," and seeks to understand what they are thinking (while recognizing that any findings are unlikely to be comprehensive). Other main concerns included: what are the issues in understanding creativity support tools, as opposed to other tools, and how do we create measurement techniques that ask whether the tool supports creativity, rather than, say, innovation or productivity?

An example of using multiple methods and metrics for study of creativity.

Creativity is a socially defined activity (Csikszentmihalyi, 1997). As such, measures of a creativity support tool's success are partially dependent on how success is defined and evaluated within a specific community of practice. Consequently, traditional measures such as performance or efficiency, while still important, are only one lens with which to view the value of a creativity support tool. To gain a more holistic perspective of how a tool influences the creative process, one may find it necessary to define new ways of measuring the impact of a creativity support tool on the problem solving process, where these metrics are derived from practices deemed important by the community under investigation.

In this section, we use research by Terry and Mynatt and their associates (2002a, 2002b; Terry, Mynatt, Nakakoji & Yamamoto, 2004) as a case example study to illustrate how a deep understanding of a community of practice can not only inform tool design, but can guide subsequent evaluations. This case study highlights the importance of employing mixed methods for evaluation, and argues that a richer suite of evaluation instruments is necessary to study creativity support tools.

Understanding User Needs

If one optimizes current practices (that is, *what* people do), one can expect only incremental improvements in computer-based support for the creative process. On the other hand, understanding *why* people do what they do can lead to new insights into the forms computational support can take. These data also indicate how one should judge the success of any computational tool support developed. Thus, before the design and evaluation of creativity support tools can commence, one should have a clear idea of the needs of a group of individuals, as well as an understanding of why these needs exist.

Among the many tools at a researcher's disposal, qualitative methods (e.g., ethnography, field studies, and the like) are particularly well-suited to gaining a deep understanding of the needs and methods of a community of practice. When performed by an experienced researcher, in-situ observations and semi-structured interviews can yield a rich set of data in a relatively short period of time (Millan, 2000), providing the information necessary for subsequent design and evaluation phases.

In studying the practices of graphic artists and designers, Terry and Mynatt (2002a) used field studies and semi-structured interviews to investigate how computational tools could better support day-to-day work processes in the visual arts. Their research uncovered one particular practice that seemed poorly supported by current interfaces: Generating and comparing multiple ways of solving a problem. Creating sets of alternative solutions is a common practice that serves many purposes when solving ill-defined problems (Terry, Mynatt, Nakakoji & Yamamoto, 2004; Newman & Landay, 2000). For example, it allows the researcher to make concrete comparisons between multiple solution possibilities. It can also force the user to push beyond current approaches to seek out more novel ways of solving a problem. However, despite its importance, Terry and Mynatt's studies found that current interfaces tend to lead to highly linear problem solving methods by virtue of offering few tools to facilitate exploration. These data led to the following hypotheses:

- Current interfaces lead to highly linear problem solving processes
- Tools that enable one to more easily explore and compare alternative scenarios would yield better solutions, faster
- Users would prefer tools that enable them to more easily explore

These hypotheses outlined the design space for computational tools, but also indicated deficiencies in evaluation instruments. In particular, concepts such as “breadth of exploration,” while intuitively understood, do not have precise definitions. Thus, measures of success relevant to this community, such as broad exploration, needed to be qualified to enable the measurement of this phenomenon.

Evaluating Exploration

The field studies of graphic artists and designers led to the design and implementation of two tools, Side Views and Parallel Pies, both for use in Photoshop-like image manipulation applications. Side Views (Terry & Mynatt, 2002b) automatically generates sets of previews for one or more commands and their parameters, allowing one to quickly generate sets of potential future states that can be compared side-by-side. Parallel Pies (Terry, Mynatt, Nakakoji & Yamamoto, 2004), on the other hand, streamlines the process of forking and creating sets of separate, standalone solution alternatives.

To understand the impact of their tools on the problem solving process, Terry and Mynatt employed three different studies: two controlled laboratory studies, and a third, think-aloud study. These studies differed by task type and the data collected to provide a more holistic understanding of how these tools affect various aspects of the creative process. We summarize each in turn.

The first study asked individuals to color-correct an image to make it match a known (visible) goal state. While highly artificial (since the goal was known), the task nonetheless had an ill-defined solution path. Furthermore, the known goal state allowed the researchers to algorithmically measure *how close* the subjects got to achieving the goal. This study's design, then, provided a baseline for how well subjects could perform when they knew exactly what they were looking for.

The second study's task more truly reflected open-ended creative work. Subjects were asked to develop color schemes for a wristwatch to make it depict the seasons of the year (winter, spring, summer, or fall). This more real-world task afforded a view of how the tools influence more creative tasks, but at the cost of being unable to perfectly assess how well goals were met (that is, there was no single “right” answer by which one could measure solution quality).

The first two studies were identical in design. Both were controlled laboratory studies employing a within-subjects design that varied experimental tool availability. All significant actions within the interface were logged and time-stamped allowing reconstruction of any state visited by a subject.

Subjects completed NASA TLX workload assessment forms after each task to help determine any potential impact the tools may have on perceived levels of cognitive load, and an exit questionnaire assessed user preferences for the tools. These data yielded a significant quantity of empirical data describing what individuals did, but not necessarily why. The third study was designed to get at this latter question.

The third study paired individuals to work collaboratively on the same wristwatch color scheme task as the second study. This design, inspired by the constructive interaction technique by Miyake (1986) has the advantage of externalizing thought processes as individuals work together on the problem. After all tasks were completed, an interview with the subjects enabled the researchers to probe specific questions that arose from analyzing the data from the first two studies.

Instruments such as the NASA TLX or the exit questionnaires are designed to collect data to answer very specific questions. However, user interface logs are raw data and require analysis. Returning to the earlier observation that exploration is highly prized in this community, the researchers developed a set of formal definitions for breadth and depth of exploration, backtracking, and “dead-ends.” These could then be applied to the log data to yield measures of how much each activity occurred throughout the studies. Custom visualizations were developed to help convey these concepts.

Several aspects of this evaluation plan are noteworthy:

- Task type was varied to understand how the tools operate under different conditions (tightly constrained tasks with well-defined solutions vs. ill-defined problems with no single, correct answer)
- Multiple types of data were collected, from quantitative to qualitative (user interface usage, workload self-assessment, tool preference, interview data). These data often complement one another, with one data set increasing the value of the other data set
- New metrics were defined to describe aspects of the problem solving process (well-defined measures for breadth and depth of exploration, backtracking, dead-ends)
- Visualizations were created to convey research results

From their studies, they confirmed that current interfaces lead to highly linear problem solving practices and that users more broadly explore when tools are available to facilitate this process. However, they also found a tendency to initially *overuse* these capabilities: Subjects sometimes spent *too much* time exploring and not enough time maturing a single solution. This finding provides a lesson that there can be cases where computational support can detrimentally affect creative processes, even though their design is well-motivated.

More importantly, this research yielded a new set of “yardsticks” for other researchers to employ when evaluating their own tools, specifically how to define breadth and depth of exploration, backtracking, and dead-ends. One of the routes to maturing research in creativity support tools is to promote more evaluations, especially those that afford comparisons between independent research results. Standardized instruments are one method of achieving this end.

The research of Terry and Mynatt is not unique, and makes use of strategies commonplace within the research community. However, with regards to evaluating creativity support tools, the following points are important to keep in mind:

- Creativity happens within a social context, and tools will partially need to be evaluated within this context

- Qualitative methods help reveal user needs, not merely practices (*why* they do what they do, not just *what* they do)
- Traditional metrics such as performance and efficiency are important to evaluating creativity support tools. However, there is a need for richer set of metrics for describing *how* tools influence the problem solving process, and whether these effects are desirable
- The creative process is composed of many smaller sub-tasks, some of which resemble less creative tasks such as optimization. It is important to analyze tools under these various environments to gain a holistic understanding of the tool's strengths and weaknesses in different milieus
- Both quantitative and qualitative data are necessary for evaluation. Empirical studies indicate significant effects, and qualitative studies aid in interpreting *why* these effects arise
- One should seek to learn the “boundaries” of the tool – what its strengths and weaknesses are, and report *both*. Successes are important, but the failures are just as important so others don't make the same mistakes

An example of multiple methods and metrics in a research program.

Research in Creativity Support Tools (CST) (George, Nunamaker, and Valacich, 1992) has involved use of multiple research methodologies. Multiple methodologies including mathematical simulation, software engineering, case study, survey, field study, lab experiment, and conceptual (subjective/argumentative) are illustrated based on an established taxonomy of MIS research methods (Vogel and Nunamaker, 1990). These variety of methodologies need to be called upon to address the multitude and multifaceted research questions that exist pertaining to creativity support tools.

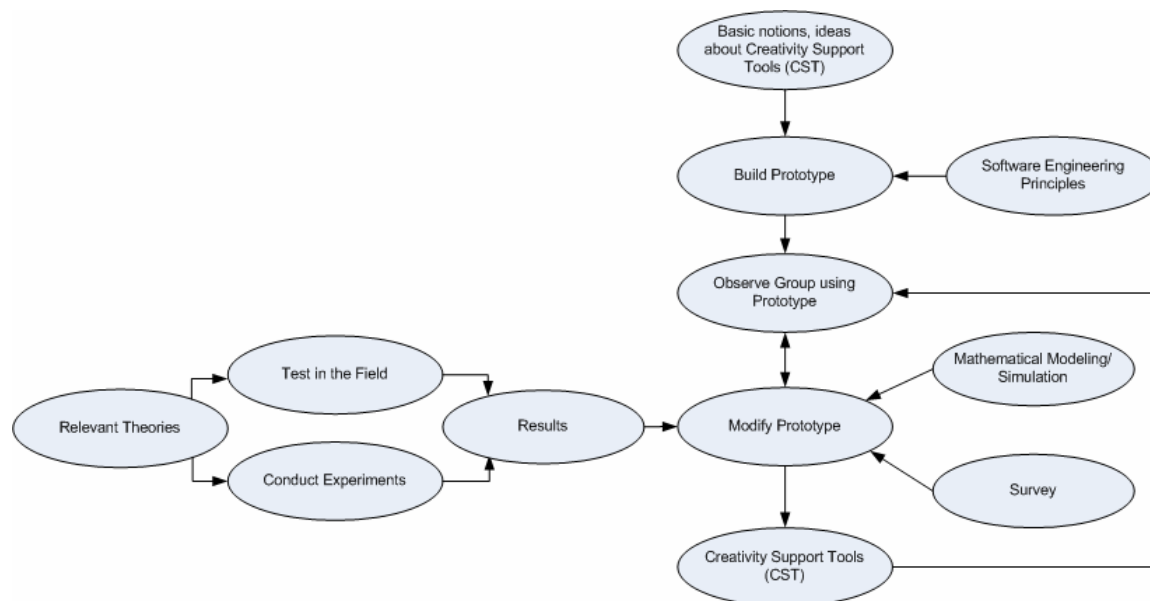


Figure 1: Creativity Support Tools (CST) Development Framework

As the Figure 1 above shows, studying the creativity process with the aid of CST involves crystallizing the few important research notions and ideas about the computer-aided creativity process by building a prototype and enhancing it with the aid of results derived from deployment of multiple methodologies. Software engineering principles are important in building the prototype. Also, early on, it is important to do preliminary testing by observing the groups using the prototype and taking the observations to improve the prototype. Mathematical modeling and simulation play a role in formalizing the process.

Methodologies such as planned experiments, case studies, and field studies are tools for empirical validation and testing that can help in improving the prototype as well as developing a theory for understanding the creativity process in the light of CST.

Creative ideas emerge from novel juxtaposition of concepts in working memory in the context of a creative task. Therefore, within the limits of human working memory, the greater the variety of concepts one considers, the greater is the probability that creative ideas will occur. Group Support Systems (GSS) helps create variety of stimuli which can be far more than those using nominal group techniques (NGT).

Many of the initial research findings related to CST came from the field settings when the initial prototypes were tested in organizational settings. Support for increasing creativity was observed in the field settings. To investigate the process from a theoretical standpoint, similar studies were conducted in the laboratory settings. Similar support was observed in the laboratory settings also. However, researchers noted certain differences between lab and field findings, many of them which can be attributed to differences in the two settings (Dennis, Nunamaker, and Vogel, 1990). Many iterations between field and laboratory experiments resulted in strong empirical validation of GSS as useful CST. Actual field use of these support tools produced effects that were not modeled or measured in the early lab experiments, often because real groups do not perform in a void, but within an organizational context that drives objectives, attitudes, and behaviors in group meetings. Our direct experience is based upon having worked with more than 200 public and private organizations in our own four meeting laboratories, as well as at over 1,500 sites around the world that have been built upon the meeting laboratory model established at Arizona. We have facilitated over 4,000 working sessions for teams and have produced more than 150 research studies in the domain of collaborative technology. An extensive review of case and field studies for CST can be found in (Fjermestad and Hiltz, 2000), while a similar review of experimental studies for CST can be found in (Fjermestad and Hiltz, 1998).

The processes and outcomes of group work in the form of creativity depend upon the interactions among four groups of variables: organizational context, group characteristics, task, and technology. What variables within these four groups have significant effects remains an open question. We have selected 24 variables; clearly there are many others that could be considered (Dennis, Nunamaker, and Vogel, 1990). However, the number of variables also precludes meta-analysis as there are too many variables for the number of studies. Our selection was motivated both by theoretical arguments (i.e. there is some a priori reason to suspect that the variable might affect processes and outcomes) and empirical findings (i.e. there have been differences among studies that have differed in these characteristics).

We discuss below two major variable groupings as related to creativity: organizational context and group characteristics. The discussion is focused around explaining the differences observed in the results from experimental and field studies of creativity tasks. In a nutshell, the results have shown that organizational groups (real groups) outperform experimental groups (nominal groups), when using creativity support tools (CST) (Valacich, Dennis, and Connolly, 1994). Only in the case of manual settings, i.e. in the case of non-CST support settings, nominal groups have been shown to outperform real groups. An extensive literature on studies with brainstorming with nominal groups can be found in (Diehl and Stroebe, 1987).

Organizational Context

We note four organizational contextual factors may be important for productivity of creative ideas. First, organizational culture and behavior norms serve as a guide to the meeting process for organizational groups in field studies. Norms may be lacking in laboratory groups formed for the purpose of an experiment; an assembled group of individuals may be a group in body, but not in spirit. In pre-existing experimental groups, contextual norms may simply be different from the norms of organizational groups.

Second, in organizational groups, group members have incentives to perform. Accomplishing the task successfully means recognition and reward for the group. In some experiments, where the tasks are such that performance can be measured objectively, this has been provided by pay and incentives based on experimental performance.

Third, members of organizational groups may not always have consistent goals and objectives; there may be political elements such that the “best” outcome for some group member(s) is different from that for other group member(s). Tasks in experiments have traditionally presumed the rational model, where organizational decisions are consequences of organizational units using information in an intended rational manner to make choices on behalf of the organization (Huber, 1981), although some (Watson, DeSanctis, and Poole, 1988) have involved what are essentially bargaining tasks which have no right answers. In field studies, objectives have not always followed the rational model. They often have a political component, where organizational decisions are consequences of the application of strategies and tactics by units seeking to influence decision processes in directions that will result in choices favorable to them (Huber, 1981).

Finally, for organizational groups, issues and problems are interrelated. Thus every time an organizational group attempts to resolve a particular problem it needs to consider the problem’s potential relationship with all other problems. This generally has not been a concern for groups in laboratory experiments.

In summary, most laboratory experiments have examined student-related organizational cultures without performance incentives but with common objectives without interrelated problems. Most field studies have examined public and/or private sectors with incentives and interrelated problems, but with necessarily the same objectives.

Group Characteristics

There have been many differences between the groups in experimental research and those in the field that may account for differences in findings. First, most experimental groups have been composed of students, while organizational groups have been composed of managers and professionals. Individual characteristics of the two populations may be different.

A second potential factor is the familiarity of group members with the task. Members of organizational groups typically have had more experience with the task, as in general, they address the tasks faced in the on-going management of the organization. In contrast, experimental groups have often had less familiarity with the task they have been assigned.

Third, experimental groups have typically been ad hoc, formed for the sole purpose of the experiment, and have no past history or foreseeable future. Field studies have typically used established groups, for whom the meeting under study is just one meeting in a long series of meetings.

Fourth, participants in experimental studies have generally been peers. While some experiments have studied the effects of an emerging leader or have temporarily assigned a leader for the duration of the experimental session (George, Easton, Nunamaker, and Northcraft, 1990), this form of leadership can be difference than that in organizations. Groups in previous field studies have generally had a distinct hierarchy and/or differing social status among members; the leader was the leader before, during, and after the meeting.

Fifth, participants in experiments have often been first time users of the CST technology. Participants in field studies have also often been first time users, but by the end of the study they have often logged many

hours of use; they are moderately experienced CST users. Observations drawn from a study of inexperienced users of CST may be useful, but may apply on to inexperienced users.

Sixth, researchers have speculated that CST may prove effective for larger groups than smaller groups (Dennis, George, Jessup, Nunamaker, and Vogel, 1988; DeSanctis and Gallupe, 1987). Most experimental research has been focused on small groups (often 3-6 members). In contrast, most field study groups have been larger (typically 10 or more members). Group size, of course, has been shown to have significant impacts on non-CST supported creativity group work.

Finally, the logical size of the group in addition to the physical size of the group is also important. Groups can be considered logically small if there is high overlap in the participants' domain knowledge and skill. The overlap or lack of overlap of skills, traits and abilities has been shown to have different effects in studies of non-CST supported creativity group work.

References

- Boden, M. (1990) *The Creative Mind: Myths and Mechanisms*, Basic Books, New York. (Precis, with peer reviews, in *Behavioral and Brain Sciences*, 17: 3, 1994.)
- Campbell, D. T. & Stanley, J. C. (1966). *Experimental and Quasi-Experimental Designs for Research*. Rand McNally, Chicago
- Campbell, D. T. & Fiske, D. W. (1959). Convergent and Discriminant Validation by the Multitrait-Mutimethod Matrix. *Psychological Bulletin*, 56, 81-105.
- Candy, L. & Edmonds, E. A., 1994. Artifacts and the designer's process: Implications for computer support to design. *J. Design Sci. and Tech.* 3, 11-31.
- Candy, L., & Edmonds, E. A., 1995. Creativity in knowledge work: A process model and requirements for support, in: *Proc. OZCHI '95*, pp. 242-248
- Candy, L. & Edmonds, E. A., 1996. Creative design of the Lotus bicycle: Implications for knowledge support systems research. *Design Studies*, 17, 71-90.
- Candy, L. & Edmonds, E. A., 1997. Supporting the creative user: A criteria based approach to interaction design. *Design Studies*, 18, 185-194
- Cook, T. D. & Campbell, D. T. (1979). *Quasi-Experimentation*. Rand McNally, Chicago.
- Csikszentmihalyi, Mihaly. *Creativity: Flow and the Psychology of Discovery and Invention*. Harper Perennial. 1997.
- Dennis, A.R., George, J.F., Jessup, L.M., Nunamaker, J.F., Jr., and Vogel, D.R. "Information Technology to Support Group Work," *MIS Quarterly*, 12, 1988, pp. 591-624
- Dennis, A.R., Nunamaker, J.F., Jr., and Vogel, D.R. "A Comparison of Laboratory and Field Research in the Study of Electronic Meeting Systems," *Journal of Management Information Systems*, 7(2), 1990-91, pp. 107-135.
- DeSanctis, G., and Gallupe, R.B. "A Foundation for the Study of Group Decision Support Systems," *Management Science*, 33, 1987, pp. 589-609.
- Diehl, M. and Stroebe, W. "Productivity Loss in Brainstorming Groups: Towards the Solution of a Riddle," *Journal of Personality and Social Psychology*, 53(3), 1987, pp. 497-509.
- Fjermestad, J., and Hiltz S.R. "Group Support Systems: A Descriptive Evaluation of Case and Field Studies," *Journal of Management Information Systems*, 17(3), 2000, pp. 113-157.
- Fjermestad, J., and Hiltz S.R. "An Assessment of Group Support Systems Experimental Research: Methodology and Results," *Journal of Management Information Systems*, 15(3), Winter 1998-99, pp. 7-149.
- Gardiner, H., 1993. *Creating Minds*. Basic Books, New York.

- George, J.F., Nunamaker, J.F., Jr., and Valacich, J.S. "Electronic Meeting Systems as Innovation: A Study of the Innovation Process," *Information and Management*, 22, 1992, pp. 187-195.
- George, J.F., Easton, G.K., Nunamaker, J.F., Jr., and Northcraft, G.B. "A Study of Collaborative Group Work With and Without Computer Based Support," *Information Systems Research*, 1(4), December 1990, pp. 394-415.
- Hewett, T. T. (1986). The role of iterative evaluation in designing systems for usability. In M. D. Harrison and A. F. Monk (Eds.), *People and computers: Designing for usability* (pp. 196-214). Cambridge: Cambridge University Press.
- Hewett, T. T. (in press). Informing the design of computer based creativity support environments. To appear in the *International Journal of Human Computer Studies*.
- Huber, G.P. "The Nature of Organizational Decision Making and the Design of Decision Support Systems," *MIS Quarterly*, 5(2), 1981, pp. 1-10.
- Mayer, R. E. (1999). Fifty Years of Creativity Research. In R. J. Sternberg (1999). *Handbook of Creativity*. Cambridge University Press. Cambridge, UK.
- Millen, David R. Rapid ethnography: time deepening strategies for HCI field research. In *Proceedings of the conference on Designing interactive systems: processes, practices, methods, and techniques* (DIS 2000). pp. 280-286.
- Miyake, N. Constructive Interaction and the Iterative Process of Understanding. *Cognitive Science*, Vol. 10, 1986, pp. 151-177.
- Nickerson, R. S. (1999). Enhancing Creativity. In R. J. Sternberg (1999). *Handbook of Creativity*. Cambridge University Press. Cambridge, UK.
- Newman, Mark W., & Landay, James A. Sitemaps, storyboards, and specifications: a sketch of Web site design practice. In *Proceedings of the conference on Designing interactive systems: processes, practices, methods, and techniques* (DIS 2000). pp. 263-274
- Sternberg, R. J. (1999). *Handbook of Creativity*. Cambridge University Press. Cambridge, UK.
- Terry, Michael & Mynatt, Elizabeth. Recognizing Creative Needs in User Interface Design. In *Proceedings of the Fourth Conference on Creativity & Cognition* (2002a). pp. 38-44
- Terry, Michael & Mynatt, Elizabeth. Side Views: Persistent, On-Demand Previews for Open-Ended Tasks. In *Proceedings of the 15th Annual ACM Symposium on User Interface Software and Technology* (UIST 2002b). pp. 71-80.
- Terry, M., Mynatt, E. D., Nakakoji, K., and Yamamoto, Y. Variation in Element and Action: Supporting Simultaneous Development of Alternative Solutions. In *Proceedings of the 2004 Conference on Human Factors in Computing Systems* (CHI 2004). pp. 711-718.
- Valacich, J.S., Dennis, A.R., and Connolly, T. "Idea Generation in Computer-Based Groups: A New Ending to an Old Story," *Organizational Behavior and Human Decision Processes*, 57, 1994, pp. 448-467.
- Vogel, D., and Nunamaker, J.F., Jr. "Group Decision Support System Impact: Multi-Methodological Exploration," *Information and Management*, 18, 1990, pp. 15-28.
- Watson, R.T., DeSanctis, G., and Poole M.S. "Using a GDSS to Facilitate Group Consensus: Some Intended and Unintended Consequences," *MIS Quarterly*, 12(3), 1988, pp. 463-478.
- Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1969). *Unobtrusive Measures*. Rand McNally, Chicago.
- Webb, E. J., Campbell, D. T., Schwartz, R. D., Sechrest, L., & Grove, J. B. (1981). *Nonreactive Measures in the Social Sciences*. Houghton Mifflin, Boston.