Critique of Comparing the Generativity of Problem Solving and Appreciative Inquiry: A Field Experiment by Bushe, G. R. and Paranjpey, N.

Paul J. Blois

Keywords
appreciative inquiry, inspiring questions, research methods, sample size, synergenesis

Foreword
This critical review of Bushe and Paranjpey’s (2015) research regarding appreciative inquiry springs from a desire to improve in asking open questions. Education Leadership and Management faculty in the Master of Arts program at Royal Roads University asked questions in such a way as to spark thinking and welcome contributions. Learning about appreciative questions (Lahaye, 2015, personal communication) confirmed this, as did analysis of the Teaching Perspectives Inventory (TPI) by Pratt and Collins (2015) and the Leadership Practices Inventory (LPI) by Kouzes and Posner (2015). Conversation regarding the LPI confirmed how effective probing questions can help develop open communication, as did the TPI area Developmental Perspective.

Synergenesis
Bushe and Paranjpey (2015) undertake a comparison of generative methods, which are essential in any organizational development transformational process (p. 310). They demonstrate in this study that the most generative method is synergenesis, a refinement and development of appreciative inquiry by Bushe since 1995 (p. 315). Synergenesis is similar to Lahaye’s (2015, personal communication) “Elements of Success” class activity, wherein we shared stories of a peak experience related to the topic, then reported our partner’s story to a group for recording. In synergenesis, according to Bushe and Paranjpey (2015), partners must also write each other’s story in the first person, and in the sharing group there must be enough time for everyone to discuss all of the ideas evoked by the stories that bear on the focal question before moving on to share the next story (p. 315).

Comparing Problem Solving with Appreciative Inquiry and Synergenesis
Bushe and Paranjpey (2015) compare the ideas generated from the brainstorming typically found during problem solving with those generated by the discovery phase of appreciative inquiry and synergenesis. The study’s methodological approach is clearly
stated and explained in enough detail for its essentials to be reproduced with just a bit of guess-work. The focus topic was the creation of an employee recognition program which could be implemented in an organization (p. 316). The quality of ideas was assessed by expert raters using a 5-point Likert scale to assess their novelty, interest, and practicality (p. 319). Researchers balanced the number of participants involved in each of the three methods according to experience and position (p. 321), and controlled the time spent in each of the methods to place them on equal footing (p. 317). They solicited for volunteers from a large organization of ten thousand employees, garnering seventy-six participants (p. 316). These participants were distributed into six sessions: two problem-solving, two classic appreciative inquiry, and two synergenes (p. 317). Each of these sessions was conducted in a session involving pair-work, and then sharing groups of five to seven (p. 317). According to my calculation, this should mean that each session involved the facilitation of two groups of five to seven, though the researchers were not clear on this point and used the words group and session interchangeably. A figure would assist in description of how a session was conducted, or further explanation to clarify this point.

The researchers used a mixed method approach, as described by Hiebert (2015, personal communication), introducing elements of complexity in a pragmatic way. Their Likert scales are a brilliant way to quantify measures of quality to describe the output of a process. They describe how a simple quantitative approach, such as simply counting the number of ideas, though common practice, is not be very useful (p. 313). The findings of the study might be limited because of the use of volunteers as opposed to randomly selected participants. Lack of a random selection process might mean that the findings cannot be generalized. Muijs (2008) explains that volunteer sampling is problematic because the people who tend to volunteer are not typical (p. 145). Bushe and Paranjpey (2015) gave careful consideration to the qualities of the raters, lending confidence to their appraisal. Raters were “blind to the experiment”, “from varied departments”, “had more than 5 years of experience with the organization”, and rated independently (p. 320).

The researchers did not claim that the ideas generated by appreciative inquiry were statistically significant in novelty, interest, or practicality compared with problem-solving, which was their Hypothesis 1. Upon inspection, they can be calculated as 4%, 5%, and 7% higher respectively, however, lending confidence to their statement that this hypothesis should be statistically significant with a larger study (p. 325). The importance of sample size in discussed in more detail in the appendix included at the end of this article, titled statistical power. Regarding the lack of statistically significant results for appreciative inquiry, Bushe and Paranjpey (2015) also admit that appreciative inquiry was not fully implemented. The dream phase, and design, added to the discovery phase of appreciative inquiry may have produced more ideas (p. 330).

**Impact of an Inspiring Question**

Their research demonstrated statistically significant ratings regarding the interest and practicality of ideas produced by synergenesis as compared with either of the other methods, confirming their Hypothesis 2 (p. 322). The differential results in favor of synergenesis might be explained because of a biased approach, however. First, in the beginning of the ideation process, facilitators along with the author selected “the most stimulating and inspiring ideas” (p. 318) to be explored further in the second part of the
session. This demonstrates the researcher’s personal involvement in the success of this process, and may have granted the synergenesis condition an attention and enthusiasm not afforded to appreciative inquiry or problem-solving. Second, in the final part of the synergenesis ideation process, participants were invited to brainstorm using an inspiring guiding question, “What types of recognition make people feel valued and appreciated, inspiring them to come to work every day and do their best?” (p. 318). In the other two conditions, on the other hand, the guiding questions were not personally involving, and simply focused on the success of an employee recognition program. For these reasons, the research results favoring synergenesis might be explained as the effect of introducing an inspiring guiding question. The researchers did not consider this, but this possibility remains an educationally informative and very promising alternative.

**Brainstorming a Problem Lowers Employee Morale**

One striking finding in the study is the “steep drop in ideas after the problem solving discussions” (p. 323). A qualitative note adds color and insight into differences in quantitative findings before and after semantic mapping: whereas the problem-solving sessions left participants frustrated and less willing to invest as time went on, participants engaged in appreciative inquiry and synergenesis finished strong, many even chose to hang around to chat afterwards (p. 326). The frustration expressed by employees while focused on an organizational problem, and their lower output, might be explained by the sensitivity of the question. Tourangeau (2008) describes personal risk as a sensitive area (p. 813). It is possible that these employees felt that their job was at risk were they to engage honestly in the process, or that expressing themselves might threaten their job security, explaining the reduced performance. This speaks to the importance of maintaining and creating positive energy as a leader among staff and within a team (Lahaye, 2015, personal communication). Bushe and Paranjpey (2015) suggest *evaluation apprehension*, the “concern that one’s ideas will be negatively judged by others” (p. 314), as another explanation for reduced performance in brainstorming.

**Concluding Remarks about the Quality of the Research**

Bushe and Paranjpey were surprised that the novelty of ideas were not significantly lower for problem-solving according to their quantitative measure (p. 326). While a 4% difference can be noted, their qualitative inquiry investigating the matter should be appreciated. Raters had an open-ended chance to explain, and as a result the researchers concluded that given the years of experience of the raters, it was not easy for ideas to surface that have not been thought of before. Raters said that some of the ideas that had been tried before might succeed again with improved implementation (p. 326). This demonstrates proficient use of mixed methodology. They took time to probe further, adding depth of understanding to their research (Hiebert, 2015, personal communication). Hiebert invites us to question how many references are made to “relevant literature in the field” when assessing the quality of research. Bushe and Paranjpey’s references are ample and adequate, though it should be noted that 10% of those listed in the reference section are to Bushe’s own previous works (p. 332). Beyond this, references
to Bushe’s previous works are used more frequently and to a greater extent in their text. This might suggest bias. That said, synergogenesis is a great contribution to the field. This research helps us ask questions and improves communication skills which will facilitate teamwork and distributed leadership.

Suggestions for Further Research

Bushe and Paranjpey provide several suggestions for future research: what might longitudinal studies show; what contributes to the generation of ideas and where they emerge; how does priming contribute to this generation; how might exploration of tactile stimulation and visual metaphors contribute to the process (p. 331). This suggestion is reminiscent the powerful and spiritually engaging rock activity conducted on our last day of residency in the Master of Arts in Education Leadership and Management program at Royal Roads University. We chose rocks that represented key learning in a reverent atmosphere, wrote one word on the rock, shared with the group, and exchanged rocks. This author suggests further investigation into the impact of inspiring questions as related to appreciative inquiry and learning and leadership in the field of Education in general.

Appendix – Statistical Power

Lavrakas (2008) explains that a null hypothesis (Ho) involves the case where the researchers will claim that there is no relationship between the variables (p. 542), in other words, the case wherein their hypothesis that there is a relationship is not true, hence null.

Buskirk (2008) defines statistical power as "probability of correctly rejecting a null hypothesis that is false" (p. 845), in other words, the chance of finding evidence for a hypothesis. Statistical power, 1 – β, ranges "from 0 to 1 with larger values being more desirable," and higher statistical power requires a greater sample size (p. 845).

The reason that sample size must be increased is because of two types of errors, both of which we want to reduce. Bobashev (2008b) explains, intuitively Type II error (β) should be reduced, however, decreasing Type II error (β) increases Type I error (α), if sample size remains the same (p. 914).

Type I error (α), Bobashev (2008a) explains, happens when the null hypothesis (Ho), referring to lack of relationship between variables, is rejected even though it is true. If alpha = 0.05, and if the test was "conducted many times where the null hypothesis is true, one can expect that 5% of the tests will produce an erroneously positive result." Type I error is also called false positive, because the test showed an association or effect when it should not have (p. 913), hence it is also called false discovery (p. 914).

Type II error (β), Bobashev (2008a) goes on to explain, is the case where the null hypothesis (Ho) is not rejected when it is actually false (p. 913). In other words, Type II error (β) refers to the case of failing to show that the hypothesis is true (p. 914).

We can consider how this applies to acceptance of the effectiveness of a school intervention. If we accept easily, then α increases while β is reduced, and vice versa.

1 Of the sixty-four references made through their text, thirteen are to Bushe’s own works. This means that references made to Bushe’s previous work comprise 20% of the references, double what should be expected if fair attention was given to all references.
References

DOI: http://dx.doi.org.ezproxy.royalroads.ca/10.4135/9781412963947

DOI: http://dx.doi.org.ezproxy.royalroads.ca/10.4135/9781412963947


DOI: http://dx.doi.org.ezproxy.royalroads.ca/10.4135/9781412963947


DOI: http://dx.doi.org.ezproxy.royalroads.ca/10.4135/9781412963947

About the Author

Paul J. Blois BSc. BEd. is a Bahá’í passionate about world unity, religion and improving education. He was a National Star Cadet Leader Instructor, international performer and national athlete. He has 3 minors in Education and Certificates in Innovative Leadership, Montessori, and the Virtues Project. He teaches in China (email: sirblois@alumni.sfu.ca).