Optimizing Design Thinking: Theatrical Improv's Affects on Creativity Within Groups

Nicholas Hennigan
Mechanical Engineering
Milwaukee School of Engineering
Milwaukee, WI, USA
hennigannm@msoe.edu

ABSTRACT

This study builds upon current research correlating the relationship among improvisation (improv), divergent thinking, and emotions. A recent study (Lewis & Lovatt, 2013)[1] looked at the affect improv had on an individual's ability to think divergently and affect their emotional state. This research extends that inquiry and examines the question: How does improv affect a group's ability to think divergently as it relates to the Design Thinking process? As an ongoing research project, design protocol has been iterative in nature. This paper aims to detail the iterative process and results of its findings. Initial protocol used a modified version of the Alternative Use Test (AUT) and traditional Profile of Mood States - Short Form (POMS-SF) to measure a groups divergent thinking ability and mood. Current design revision uses the same Modified Alternative Use Test (AUT) coupled with an observational metric for group behavior in lieu of the POMS-SF. Participants were grouped into cohorts of 4 and sorted into either: Treatment" or "Control" groups.

CCS CONCEPTS

• Social and professional topics → Professional topics

KEYWORDS

divergent thinking, design thinking, improvisation, group dynamics

ACM Reference format:

Nicholas Hennigan. 2020. Optimizing Design Thinking: Theatrical Improv's Affect on Creativity within Groups. In ACM Grand Final's 37th ACM Global Student Research Competition

1 INTRODUCTION

While the field of Technical Communication has only started investigating the value of introducing theatrical improvisation (improv) in the undergraduate classroom [2], evidence suggests the benefits of these theater-inspired games can build collaboration skills in participants [3] [4]. This paper focuses on the relationship between improv and design thinking. Specifically, improv's impacts on idea generation within groups.

Two core components of the design thinking methodology are empathy (human centered design) and "ideation" (ability to generate many ideas) [5]. Together, these components are defined as divergent thinking [6]. However, because of its complexity, divergent thinking can be difficult to implement and measure in academic or professional settings. In fact, recent work with the University Innovation Fellows program at Stanford University's Hasso Plattner Institute of Design [7] modified the design thinking process by incorporating improv exercises to improve group empathy and ideation.

As a University Innovation Fellow at Milwaukee School of Engineering, Nicholas Hennigan is looking to expand on a recent study [1] that correlates improv with proficiency in idea generation. Specifically, exploring quantitatively improv's affect on a group's ability to think divergently. As an ongoing study, the procedure has evolved with each iteration of the study. This paper seeks to synthesize the findings of these iterations. In general, the core hypothesis has not changed - That a group develops familiarity and collaboration skills through the extended interaction via improv and will score higher (compared to a control) on two tests intended to measure divergent thinking and brainstorming effectiveness. What has changed is the evaluations used to measure divergent thinking and brainstorming.

1.0.1 Initial Design Tools: The first iteration of the study exactly copied the tools used in Lewis and Lovatt's paper [1] to build off previous works. However, the emotional survey (POMS-SF) was eventually found to not fully capture the effect of brainstorming.

1.0.2 Current Design Tools: While keeping the metric for divergent thinking unchanged, the POMS was replaced with an observational tool to better describe group interaction.

2 STUDY DESIGN

2.1 Initial Design Procedure

Between each study iteration, design procedure structure remained relatively unchanged and comprised of a "pre-", "Treatment,'' and "post-" phase in which a control and treatment group was evaluated independently and then compared. This design follows closely to the procedure published by Lewis and Lovatt [1]. Key differences are the focus on group behavior, improv exercise selection and 25% reduction in exposure to improvisation.

In the "pre-" phase of the experiment, participants were registered and assigned an ID number along with an initial POMS-

SF emotional evaluation to assess their current mood. The purpose of the POMS-SF is to observe any emotional changes among participants when subjected to the treatment. Although Lewis and Lovatt [1] found no correlation between mood and increased divergent thinking ability, research points to potential emotional benefits when looking at group improv interactions [4]. After registration, ID numbers are randomized and sorted into a cohort of four. Cohorts would further be randomly sorted into either the control or treatment groups. Once allocated to a room, an initial AUT test was given. To reduce any formation of bias between retesting, AUT evaluation were counterbalanced where two versions of the evaluation were given. The control group was provided version A while the treatment group provided Version B. Due to the POMS-SF evaluating a participant's current emotional state, counterbalancing was not needed.

In the "treatment" phase, the treatment group was exposed to 15 mins of improv exercises derived from works of Spolin [8] and Johnstone & Wardle [9]. The control group was exposed to 15 minutes of scripted verbal interaction. The scripted verbal interaction was done to mimic the social exposure experienced in the treatment group while filtering out the cognitive exercise component of theatrical improv. After exposure to improv or scripted verbal interaction, a "post-" phase was initiated. The control group was administered AUT - version B with the treatment group given version A. After administration of the second AUT, a final POMS-SF was distributed to all participants and collected. This experimental procedure can be seen graphically in Figure 1.

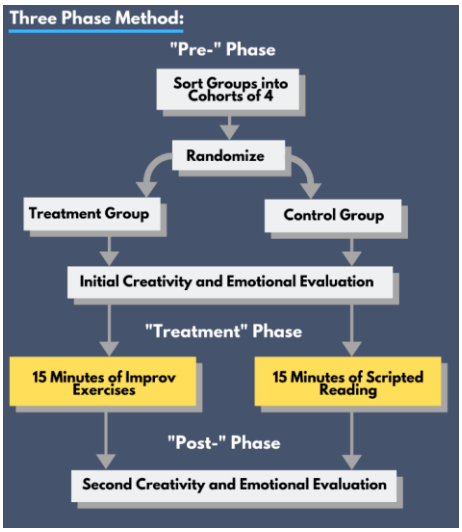


Figure 1: Initial Design Procedure for Evaluating Brainstorming and Divergent Thinking

2.2 Updated Design Procedure

The updated procedure retained the same general structure of the "pre-", "Treatment,'' and "post-" phase in which a control and treatment group was evaluated independently and then compared. Differences are noted in participant randomization and removal of the POMS-SF emotional survey.

2.2.1 Randomization Procedure

The group randomization protocol was analyzed and deemed an important characteristic to optimize. When evaluating cohort brainstorming performance (Divergent Thinking) the proficiency level of each individual could be a factor in the group's ability to perform well. With that assumption, randomizing groups does not provide enough control on whether those groups contain an even distribution of participants who are both experienced and unexperienced at brainstorming. To accommodate this factor, a Qualtrics [10] survey was created with a brief questionnaire of 3 questions to determine whether a participant was proficient with brainstorming, improv exercises and/or group work. An example question can be seen in Figure 2.

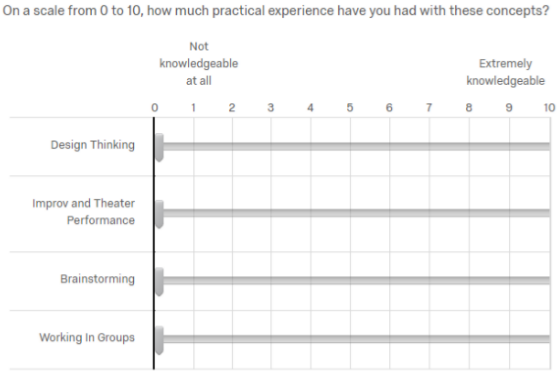


Figure 2: Example of One Question Form Qualtrics Survey

Once the surveys are collected, they populate an excel spreadsheet. Each participant is generated a "Comfort" score described by equation 1. The sum of their point values divided by the total amount of point possible determines a participants comfort score. As an example, if a participant rated themselves as being 8 out of 10 for experience in design thinking, the portion of their Comfort score for design thinking would be 0.8. This value would be added to their rating for Improv, Brainstorming, etc to determine their overall score.

$$Comfort\ Score = \frac{\sum Participants\ Scores}{Total\ Possible\ Points\ Avalible} \quad (1)$$

Once each participant has generated a Comfort score, all the scores are averaged to produce a Mean Comfort Score (MCS) for the pool of participants. Finally, a MATLAB sorting algorithm was created to generate groups of 4 that met the Mean Comfort Score within one standard deviation. This produces a best possible selection of randomized participants that are also uniform in their brainstorming experience.

2.2.2 Observational Collaboration Metric

The second alteration to the procedure was replacing the POM-SF emotional survey with an observational metric from a study titled *"Observing Collaborative Problem-Solving Process and Outcomes"* by Wilczenski et al (2001) [11]. This was decided based on results and observations of the first iteration of the study design. Noticeable behavior changes were observed in the initial

study that neither the AUT nor POMS captured appropriately. The AUT provided insight into divergent thinking ability and POMS looked at individual's emotional response. However, notable social changes were seen. This is further discussed in Section 4. As a result, the observational metric created by Wilczenski et al (2001) [11] was introduced to the procedure. This final design alteration can be seen graphically in Figure 3.

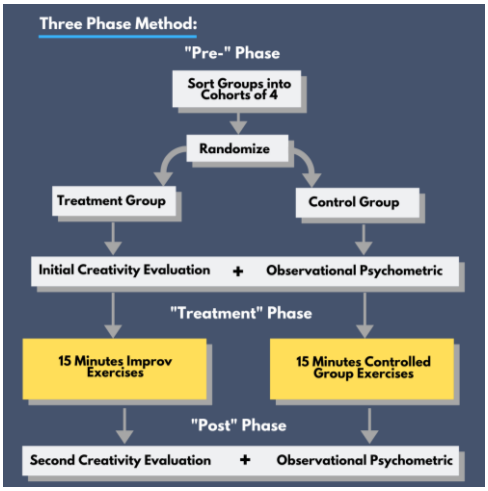


Figure 3: Current Initial Design Procedure for Evaluating Brainstorming and Divergent Thinking

Due to the addition of an observational metric, inter-rater reliability became an additional variable. To address this added complexity, the study was changed to a muli-site study by involving Cardinal Stritch University's Clinical Psychology program [12]. By incorporating Cardinal Stritch, a graduate student in the clinical psychology program was added as an additional investigator. With a second investigator, interrater reliability could be better maintained.

3 MEASURES

All measures are described and listed alphabetically. Each metric is a common psychological evaluation or modification of a previously published psychometric tool. Each measure listed was used in either the first iteration or current design procedure to test the hypothesis. The measures used through the research process are as follows: A modified version of the Alternative Use Test, Observational Collaboration Metric created by Wilczenski et al (2001)[11], and the traditional *Profile of Mood States - Short Form*.

3.1 Alternative Use Test (AUT)

The Alternative Use Test (AUT) is an evaluation designed to measure divergent thinking—a parameter related to an individual's ability to be creative [13][14]. Participants are given an ordinary object (i.e paper clip) and asked to list as many alternative uses for said object. Participants are then rated on four criteria: originality (how unique the answer is to the total population), fluency (number of generated responses), flexibility (how many "categories" each "use" falls into), and elaboration (amount of detail provided for each answer). In order to quantify a group's ability to think divergently, the AUT was modified by being administered to an entire group instead of individuals. Although this modification is novel to the study, the ability to evaluate a groups creative performance is necessary for the study and warrants potential deviation from the published form. Anecdotal evidence from facilitating over 21 Design Thinking workshops reaching 400 students suggests a groups creative potential appears not to be the sum of each individual's creative ability. It seems, individuals express reticence when brainstorming within a group compared to working alone. The anecdotes suggest the divergent thinking ability of the group should be considered holistically. As such, the AUT criteria was scored to evaluate the group's performance as an ensemble.

3.2 Profile of Mood States (POMS-SF)

Profile of Mood States-Short Form (POMS-SF) [15] is an abbreviated psychological mood rating scale comprised of 37 questions (compared to the traditional 65) [16] to measure a participant's current emotional state. Administration takes 5-15 minutes. Participants are prompted with a single word and asked to rate how relevant it is to their current mood. This rating system uses a 5-point response scale ranging from "not at all" to "extremely." The evaluation measures six different dimensions of mood swings: Tension/Anxiety, Anger/Hostility, Vigor/Activity, Fatigue/Inertia, Depression/Dejection, Confusion/Bewilderment. Although research done by Lewis and Lovatt [1] showed no emotional change with improv exposure on individuals, the test will be replicated to determine if prolonged group exposure alters emotional states.

3.3 Group Collaboration Metric

A publication titled "Observing Collaborative Problem-Solving Process and Outcomes" by Wilczenski et al (2001) [11], looked at how groups of 4 students worked collaboratively on math problems. The collaboration tool aimed to evaluate individual's capacity for collaboration as well as the behavior of the group. This focus of group behavior is commonly described as the study of group process.

According to Wilczenski et al (2001) [11], group process was scored and observed by 3 independent observers based on facilitating and detracting behaviors. Examples of scoring included behaviors that facilitated group collaboration (e.g. facilitating conversation, asking clarifying questions and applying a strategy) and detracting collaboration (e.g. not paying attention, interrupting, and monopolizing). These behaviors were scored for each student and the groups collectively. To address repeatability, the sessions were videotaped to increase the number of observers and a general group score was determined form the aggregate.

Reliability:
From the pool of observers, coefficients of agreement were determined by the following model seen in equation 2 (Sattler, 1992) [17].

Additionally, to address any behaviors not recorded that may inflate scores, the random chance model of Hopkins and Hermann (1977) [18] generated a "greater-than-chance lower boundary by which acceptable interobserver agreements could be judged" (Wilczenski et al, 2001) [11]. This produced a reliability score of 0.80. The observations for detracting behaviors were rare in most groups and due to the low volume of data, a reliability score was not generated due to statistical insignificants of the few observations.

$$C_{observer} = \frac{\#\ of\ Agreements\ for\ an\ Observer}{\#\ of\ Agreements + \#\ of\ Disagreements} \quad (2)$$

Modifications
From the initial research method, observations in group collaboration were noticed. These observations were not accounted for with the AUT evaluation alone. Using the metric devised by Wilczenski et al (2001) [11], a measurement of group collaboration was achievable. Alterations to the metric to fit the study was done in two parts.

(1) Observation Technique: This experiment will not be externally recorded via film or audio. A facilitator and secondary observer will be in the room to evaluate the group observations. This reduction in observer number reduces the capacity of parameters this study can record. To address this, the portion of the evaluating an individual's behavior will be omitted. Instead, the second half (group observation) will be the primary metric and be scored the same (i.e. the general group will be evaluated for how many behaviors are noticed in the session).

(2) Behaviors Being Observed: An example list of scored behaviors published by Wilczenski [11] can be seen in in Table 1. In order to adjust the tool for our subject of study (brainstorming instead of a math problems) and age range (College students vs Elementary students) the listed behaviors will be modified to more accurately represent our data sample without deviating from the intent of published behavioral indicators. The modified list of behaviors can be seen in Tables 1 & 2.

Table 1: Standard Behavior Indicators from Published Study [11]

Behaviors That Facilitate	Behaviors That Detract
Communicating a Strategy	Lack of Attention/Distracting others
Correctly Applying a Strategy	Incorrect Application of Assumptions
Recognizing Errors	Monopolizing
Bring Others into Discussion	Personal Attack
Asking Clarifying Questions	Not Contributing to Group Discussion
Moving Discussion Along	Interrupting

Table 2: Modified Behavioral Indicators to Fit Creativity Study

Behaviors That Facilitate	Behaviors That Detract
Communicating	Lack of Attention/Distracting others
Building off an idea	Believing their answer is the best one
Facilitate through a discourse	Monopolizing
Bring Others into Discussion	Judging an idea as bad/silly/impractical
Asking Clarifying Questions	Not Contributing to Group Discussion
Writing down ideas	Interrupting

4 RESULTS

4.1 Initial Study Results

Out of the 45 participants recruited, only 2 attended the study. Although the data cannot be corelated to the hypothesis, observations of the experimental process can be made. The two participants formed a single group and were put through the improv treatment group and evaluated. Scores for the AUT can be seen in Figure 4.

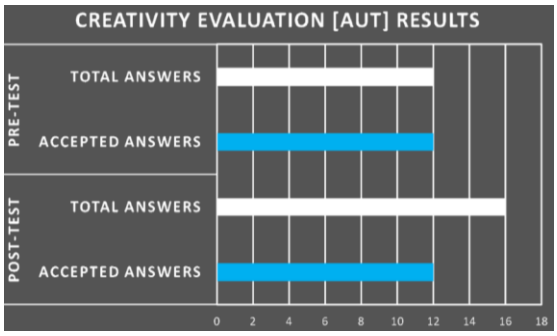


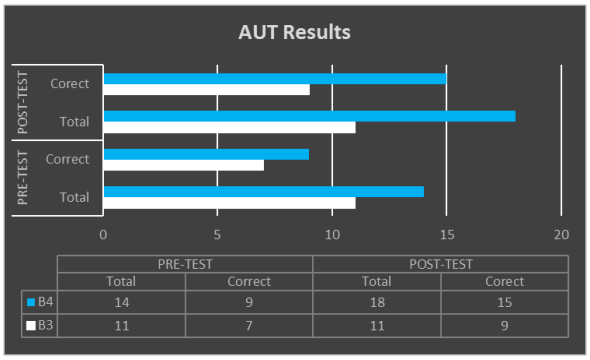
Figure 4: Comparing Number of Total & Accepted Answers for AUT Before and After Exposure to 15 Minutes of Improv

Although there is a noticeable increase in total answers between "pre-" & "post-" test, the amount of accepted answers were identical. However, group behavior was starkly different between each AUT evaluation.

The two participants were roommates and had a very strong rapport with one another. However, during the "pre-" test, they completely disassociated from one another. There was no talking and each participant was silently working on the evaluation. After improv exposure, the "post-" test produced much more social interaction. Both participants were communicative, collaborative and discussing answers to the AUT evaluation. This drastic change in group behavior was the primary reason to revise the testing methodology. None of the tools (AUT or POMS) accounted for group behavior.

4.2 Current Study Results

Out of the 150 students recruited, only 2 showed up on different days. For each evaluation, only a single participant participated in the study. Recruitment retention was most likely low due to recruiting 3 weeks before final exams. However, given the low volume of participants, results are seen in Figure 5. Figure 5 shows a comparison between participant B4 and participant B3.



	PRE-TEST		POST-TEST	
	Total	Correct	Total	Corect
B4	14	9	18	15
B3	11	7	11	9

Figure 5: Number of Total & Accepted Answers for AUT Before and After Exposure to 15 Minutes of Improv

Participant B3 produced the same total number of ideas before and after exposure. However, B3 increased their number of correct answers by 28%. Participant B4 increased their total number of ideas by 28% and had an increase in correct answers of 67%.

Interesting to note that B3 had exceptionally unique answers to the AUT evaluation. Examples of these unique answer are shown in Table 3.

Table 3: Unique Answers from Participant B3

Object	Alternative Use Submitted by B3
Key	Open portal to Gig (PlaneScape)
Shoe	Bribe the Hype-beast
Tire	Fort for Dwarf Person

These answers were so unique, the AUT manual did not classify any of them as correct or incorrect. Answers were a combination of references to specific video game lore (ex. Opening a portal in the game PlaneScape) or mythical in nature (ex. Fort for a Dwarf person & Bribing the Hype-best). The AUT has scores for Category and Uniqueness, and given these potential answers, a method must be created to deal with extremely responses.

5 DISUCSSION

Given the low number of participants, the observational metric was unable to be used. Further continuation of the study in the coming semester will provide more opportunities for increased participation. Given the advent of COVID-19, methods of facilitating sociological and psychological research digitally, should be explored.

The discovery of extremely unique answers to the AUT seen in Table 3, provide an interesting discussion for classifying what constitutes a "correct" answer. Additionally, given the Uniqueness modifier weighing the top 5% of original responses, how can one score responses that are all inherently unique?

The current solution is to generalize the response into a known category. However by doing so the observer loses resolution to detect these highly unique entries. This may be solved with larger sample sizes. As the volume of samples increase, a larger distribution of "normal" solutions will occur and bring validity to the Uniqueness Modifier.

ACKNOWLEDGMENTS

Individuals: Kristin Shebesta, M.A. (Rader School of Business, Milwaukee School of Engineering); Daniela Guerrero (Masters Student – Clinical Psychology, Cardinal Stritch University); Trevor F. Hyde, PhD (Clinical Psychology, Cardinal Stritch University); Tammy Rice-Bailey, PhD (Humanities, Social Science and Communications, Milwaukee School of Engineering); Jeffrey Rolland, PhD (Department of Mathematics, Milwaukee School of Engineering); Christine Larson, PhD (Department of Psychology, University of Wisconsin – Milwaukee); Leticia Britos Cavagnaro, PhD (Co-Director, University Innovation Fellows Program, and Adjunct Professor Hasso Plattner Institute of Design (d.school), Stanford University); Humera Fasihuddin (Co-Director, University Innovation Fellows Program, Hasso Plattner Institute of Design (d.school), Stanford University)

Institutions: Milwaukee School of Engineering; Stanford University, Hasso Plattner Institute of Design; Cardinal Stritch University; University of Wisconsin Milwaukee, Neuroscience Research

REFERENCES

[1] Lewis, C., & Lovatt, P. J. (2013). Breaking away from set patterns of thinking: Improvisation and divergent thinking. Thinking Skills and Creativity, 9, 46-58. doi:10.1016/j.tsc.2013.03.001.

[2] Rice-Bailey, T. (2019). Using Improvisational Games to Teach Collaboration Skills to University Students. Manuscript submitted for publication.

[3] Benjamin, S., & Kline, C. (2019). How to yes-and: Using improvisational games to improv(e) communication, listening, and collaboration techniques in tourism and hospitality education. Journal of Hospitality, Leisure, Sport & Tourism Education, 24, 130-142. doi:10.1016/j.jhlste.2019.02.002.

[4] Stewart, C. (2016). Effects of Improv Comedy on College Students. doi:10.30707/etd2016.stewart.c.

[5] Plattner, H., Meinel, C., & Leifer, L. (2013). Design Thinking Understand - Improve - Apply. Berlin: Springer Berlin.

[6] Runco, M. A., & Acar, S. (2012). Divergent Thinking as an Indicator of Creative Potential. Creativity Research Journal, 24 (1), 66-75. doi:10.1080/10400419.2012.652929.

[7] Zhou, L. (2018, February 05). University Innovation. Retrieved from https://dschool.stanford.edu/programs/university-innovation.

[8] Spolin, V. (2017). Improvisation for the theater: A handbook of teaching and directing techniques. Eastford, CT: Martino Fine Books.

[9] Johnstone, K., & Wardle, I. (2019). Impro: Improvisation and the theatre. London: Bloomsbury Academic.

[10] Qualtrics XM - Experience Management Software. (2020, February 26). Retrieved from https://www.qualtrics.com/

[11] Wilczenski, F. L., Bontrager, T., Ventrone, P., & Correia, M. (2001, April 6). Observing collaborative problem-solving processes and outcomes. Retrieved from

[12] Cardinal Stritch University. (n.d.). Master of Arts in Clinical Psychology. Retrieved from https://www.stritch.edu/Programs/Master-of-Arts-in-Clinical-Psychology

[13] Alternate Uses (Guilford's Alternate Uses). (n.d.). Retrieved from https://www.mindgarden.com/67-alternate-uses

[14] Mumford, M. D. (2012). Handbook of organizational creativity. London: Elsevier/Academic Press.

[15] Shacham, S. (1983). A Shortened Version of the Profile of Mood States. Journal of Personality Assessment, 47 (3), 305-306. doi:10.1207/s15327752jpa4703_14

[16] McNair, D. M., Lorr, M., & Droppleman, L. F. (1992). Edits manual for the Profile of mood states. San Diego, CA: Educational and Industrial Testing Service.

[17] Sattler, J. (1992). Assessment of Children (Rev. 3rd ed.). San Diego, CA: Jerome M. Sattler.
https://onlinelibrary.wiley.com/doi/abs/10.1002/pits.1017

[18] Hopkins, B.L., & Hermann, J.A. (1977). Evaluating interobserver reliability. Journal of Applied Behavior Analysis, 10, 121–126.