

Further Research on the Scantron Testing Method

Adam Wieczorek

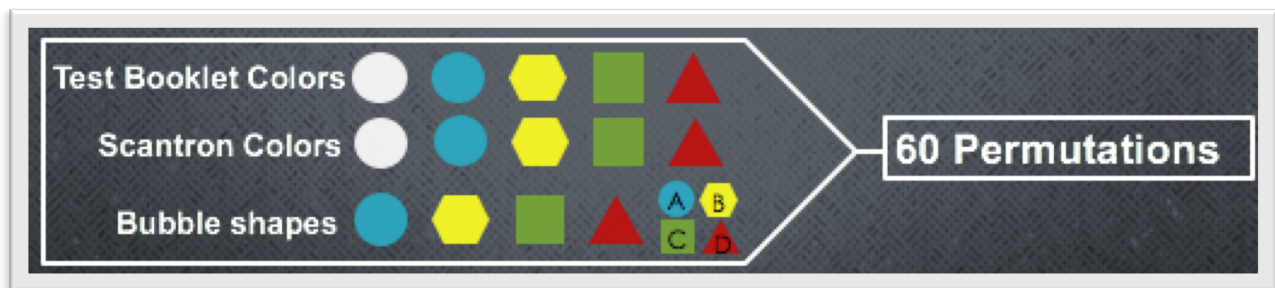
There has been little to no extensive research on the Scantron testing method. This method of testing is accepted at a rate of 98% in United States school districts (Scantron Corporation, 2017). It is widely accepted and frequently used by college universities. This method of testing is quick to grade, easy to administer, and has made the ability for students to cheat much harder. The Scantron testing method involves a sheet with fill in bubble shapes that are usually labeled with the letters a through e. This sheet is referred to as the Scantron. A test booklet that has a multitude of questions in numerical order. The questions have multiple choice answers that correspond to a letter. The letters being usually a through e. Students receive a Scantron and a test booklet that can be one of many different colors. The different colored test booklets have the same questions but in a different order. A student that sits right next to another student will not have the same colored test booklet. This nearly eliminates the ability for students to cheat by looking at another students' test booklets. Question 1 on the test booklet will correspond to question 1 on the Scantron. If the answer to question 1 in the test booklet is a then the student would fill in the bubbled labeled a on question 1 in the Scantron. The answers are then graded by a machine (Tyre, 2003). The test score results from this testing method can have a tremendous impact on a multitude of individuals. The score result from one test can have an immediate impact or can build up over time. The various people that can be impacted by these results are the individuals taking the tests, the professors administering the tests and teaching the course, and the department in which the course is in. There can be numerous different impacts from test score results. Some impacts being potentially losing or gaining a career, financial gain or loss, or being missed or recognized as a great institution. Because of these potential impacts, the wide acceptance, wide usage of this method, and

because this method has been around since the early 1970's it is shocking that this method of testing has been allowed to go on without extensive and credible research (Deffenbacher & Deitz, 1978). Since this testing method seems to be the method for testing students in universities for the foreseeable future it is important that this method is as best designed as possible and that it provides an equal experience for every student.

The theory behind this research study is that different colored test booklets combined with different colored Scantrons that have different fill in shape bubbles will produce different test score results on the same material and questions. Memory is based on the ability to associate things and recall them. These can be complex things but they can also be simple things like shapes and colors (Heil, Rösler, & Hennighausen, 1994). Shapes and colors are something we see and associate with daily (Parra, 2014). This potentially poses a problem for the current Scantron testing method. For example, an individual may associate yellow with the feeling of anxiety or they may associate that to an event in their life that causes them to feel anxious. An individual may associate the shape of a square with something that makes them feel angry. Being in any other state than being focused while taking a test has a potential for decreasing test score result. Not only do shapes and colors have different impacts on an individuals' memory, association, and emotional state, but colors and shapes also elicit different cognitive and behavioral effects by just seeing them. This is referred to as color theory (Walters, 1982). The colors commonly used for test booklets are yellow, blue, red, green, and white. Each of these colors produces a different effect by viewing them. Yellow often causes an individual to experience a state of anxiety and grabs attention. Blue promotes creativity and alertness. Red impairs problem solving and creativity. Green reduces stress and mental fatigue

while promoting problem solving and creativity. White is associated with goodness and security (Lidwell, Holden, & Butler, 2015). With this in mind a research design can be constructed.

As previously stated the different colors to be used for the Scantrons and test booklets, due to them normally being the primary colors used, will be yellow, blue, red, green, and white. The different fill in bubble shapes on the Scantrons will be either circles, squares, triangles, decagons, or a combination of the four that correspond to the letters on the test booklets. For example, circle would represent the letter a, decagon would represent the letter b, square would represent the letter c, and triangle would represent the letter d. These shapes were chosen because they are easy to fill in and are easy to administer on the Scantrons. The graph below shows the different possible combinations.



All these possible different combinations are equal to a total of 60 permutations. This information provides what sample size is needed to get enough data to accurately cover each permutation. The larger the sample size the more concrete the findings will be. However, the sample size must be practical so the experiment has the best chance to be successful. The focused population for this study is college students. College students are the optimal population for this study since they are the group that is most frequently exposed to the Scantron testing method and they are also the group this method impacts the most.

There will be 10 universities that will be selected across the United States. The universities will be randomly selected from a set of requirements. The requirements will be a large population of students and an education level that is representative of the national average for colleges. Because this research design is focusing on students' ability to answer questions on a test it is important that the education level for each university is as similar as possible and representative of the college student population. There will be 240 students that will be randomly selected from each university to give a total of 2,400 students to participate in this experiment. The remaining universities that are not eliminated by the set of requirements will be randomly selected by giving a number to each university and then putting those numbers into a random number generator. The students will be randomly selected by using their dot number in a random number generator. This method will provide a good coverage for the focused population. Before participating in the experiment students will be required to take a basic background and consent survey. The only requirement students must meet, other than giving their consent, to participate in the experiment is to not be color blind. Not being able to see the colors in the experiment defeats the purpose of color theory and this would be a threat to validity. The other reason for the survey is to identify possible patterns that may be present with the different permutations. For example, maybe females do significantly better with one combination than males do. Maybe one ethnicity, age, etc. does better with one combination. Neither the researchers nor the students will know the intent of the experiment, making this a double-blind study. This will help eliminate the experimenter expectancy and demand. The researchers and students will be debriefed on the intention and results of the experiment via email after the conclusion of the study. The experiment will take place on the same day across

the ten universities and will last only one day from 8am to 8pm. Since this experiment will only last one day there will be nearly no concern to the threats of validity due to maturation and history. The same lecture hall will be used at each university for the entirety of the experiment. It is important that the lecture halls across the different universities are as similar as possible to help eliminate the reactivity to the experimental situation and the interaction of causal relationship with settings. The lecture halls for the experiment will need to be large because each test will be the same. This will ensure that students will not sit next to anyone and thus will not have the opportunity to look at another students' tests. Each testing session will last 1 hour. Since the experiment lasts from 8am to 8pm this will give a total of 12 experiment sessions. There will be 20 students randomly assigned to each session. Because students will be randomly assigned to vastly different time slots during the day from 8am to 8pm there must be an incentive for them to attend the experiment. This incentive will either be the opportunity to win an amazon gift card or receive extra credit points for a single class. The test the students will take will consist of 20 multiple choice questions. The answers will range from the letters a through d and there will be only one correct answer per question. The questions will come from four common core subjects that are taught throughout the school system. Those subjects being math, science, english, and history. There will be 5 questions for each subject. Those questions will not be in order in regards to subject, but will be universally randomly put together. These questions are not meant to be extremely difficult or extremely easy to answer. They will be general knowledge questions that require the student to actively think about each answer, but it should be within every college students' ability to somewhat confidently answer each question. Since each session last one hour and there are only 20 questions this gives students

ample time to complete the test and should allow for researchers to start collecting and organizing the results. Students will show up to their scheduled testing time and will be able to choose where they sit so long as they do not sit next to another student. This testing experience will be exactly like a normal college testing experience, meaning no distractions and only number two pencils are to be used on the Scantron to mark their answers. Each student will be randomly assigned a Scantron color with different fill in shape bubbles, and a colored test booklet. This will be done by applying a random number to each Scantron and test booklet and then randomly assigning them.

The Scantron testing method involves viewing questions and possible answers on a variously colored test booklet and then choosing the answers by shading in a circle on a various colored Scantron. Memory is a system based on recall and association (Heil et al., 1994). People consistently view different shapes and colors daily. Because of this individuals will experience different mental, emotional, and behavioral states when they view different colors and shapes (Liu, 2016). This can potentially create an unequal testing experience because each student does not receive the same colored Scantron and test booklet. In addition, viewing different shapes and colors also generates different cognitive and behavioral effects (Lidwell et al., 2015). One result from the experiment that will support this hypothesis is variance in test score results on separate permutations. Another result that will support this hypothesis is variance in answers on individual questions or the core subjects on separate permutations. If the expected results become the actual results, there will be major real world implications for the entire college atmosphere and any other atmosphere that uses the Scantron testing method. Since 98% of United States school districts employ the Scantron testing method there must be an

urgency and immediacy to follow this study with replication studies to ensure that the results are not a fluke (Scantron Corporation, 2017). There should also be follow up studies that use different Scantron colors, fill in bubble shapes, and test booklet colors that are different from this research design. This will ensure that we can discover what permutation should be used for the Scantron testing method that generates the most equal testing experience while providing the highest test score results for every college student.

References

- Deffenbacher, J. L., & Deitz, S. R. (1978). Effects of test anxiety on performance, worry, and emotionality in naturally occurring exams. *Psychology In The Schools*, 15(3), 446-450.
- Heil, M., Rösler, F., & Hennighausen, E. (1994). Dynamics of Activation in Long-Term Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(1), 185-200.
- Lidwell, W., Holden, K., & Butler, J. (2015). *The pocket universal principles of design: 150 essential tools for architects, artists, designers, developers, engineers, inventors, and makers*. Beverly, MA: Rockport.
- Liu, S. (2016). *Color research and application: Hierarchical emotional color theme extraction* John Wiley & Sons Inc.
- Parra, M. A. (2014). *Neuropsychologia: Neural correlates of shape-color binding in visual working memory*. Elsevier.
- Scantron Corporation. (2017). Retrieved April 25, 2017, from https://en.wikipedia.org/wiki/Scantron_Corporation
- Tyre, T. (2003). Scantron Delivers Hard Data the Easy Way. *THE Journal*, 30(10), 15.
- Walters, J. (1982). *Motivation and emotion: Color preference, arousal, and the theory of psychological reversal* Springer.