Analytics in Baseball:

Retention of Sport Specific Analytic Information Based on Various Presentation Methods

by

Andrew Gin

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved November 2020 by the
Graduate Supervisory Committee:

Robert Gray, Chair
Nancy Cooke
Scotty Craig

ARIZONA STATE UNIVERSITY

December 2020

ABSTRACT

Analytics are being collected on a day to day basis on just about anything that you can think of. Sports is one of the recent fields that has started implementing the tool into their game. Analytics can be described as an abundance of statistical information that show situational tendencies of other teams and players. It is hypothesized that analytics provide anticipatory information that allows athletes to know what is coming; therefore, allowing them to perform better in real game scenarios. However, it is unclear how this information should be presented to athletes and whether athletes can actually retain the abundance of information given to them. Two different types of presentation methods (Numeric and Numeric plus Graph) and two different amounts of analytic information (High and Low) were compared for baseball players in an online based baseball specific retention survey: High Numeric (excess information shown in spreadsheet format), Low Numeric (key information shown in spreadsheet format), High Numeric plus Graph (excess information shown as a spreadsheet with hitting zone maps), and Low Numeric plus Graph (key information shown as a spreadsheet with hitting zone maps). Athletes produced different retention scores for the type of presentation method given across the whole study. Athletes presented analytic as Numeric plus Graph performed better than athletes in just Numeric condition. Additionally, playing experience had a significant effect on an athlete's ability to retain analytic information. Athletes with 10 plus years of baseball experience performed better in every condition other than High Numeric plus Graph compared to athletes with less than 10 years of experience. Amount and experience also had an interaction effect that produced statistical significance; those with less experience performed better in conditions with less baseball information given

whereas those with more experience were able to handle more baseball information at once. Providing analytic information gives athletes, especially baseball batters, a significant advantage over their opponent; however, ability to retain analytic information depends on how the information is presented and to whom the information is being presented.

TABLE OF CONTENTS

Page

LIST OF TABLES v

LIST OF FIGURES vi

CHAPTER

1 INTRODUCTION 1

The Problem 1

Background 2

Overview 5

2 METHODOLOGY 7

Design 7

Participants 7

Materials 8

Experimental Procedure 17

3 RESULTS 19

Retention Effects 19

4 DISCUSSION 24

Practical Implications 27

Limitations 28

Conclusions 30

REFERENCES 32

APPENDIX 35

A IRB EXEXPTION FOR HUMAN TESTING 36

LIST OF TABLES

Table Page

1. Means and Standard Deviations by Experimental Condition 20-21

LIST OF FIGURES

Figure Page

1. Tutorial Example Analytic Graph 9
2. Tutorial Example Questions 10
3. Analytic Information on Pitcher D 11
4. High Analytic Information Presented in Numeric Condition 12
5. Low Analytic Information Presented in Numeric Condition.....13
6. High Analytic Information Presented in Numeric/Graph Condition 14
7. Low Analytic Information Presented in Numeric/Graph Condition.....15
8. Retention Test 16
9. End Survey 17
10. Mean Retention Scores for the Different Methods of Presenting Analytics by Level of Experience. Error bars are Standard Errors 19
11. Mean Retention Scores Based on Type of Presentation Method Given. Error bars are Standard Errors 22
12. Mean Retention Scores Based on Baseball Experience Level in Years. Error bars are Standard Errors 22
13. Mean Retention Scores Based on Amount of Information Given and Level of Baseball Experience. Error bars are Standard Errors 23

CHAPTER 1

INTRODUCTION

The Problem

Baseball is a sport driven by probabilities and statistics: batting average, fielding percentage, strike rates, etc. (Baumer, 2014). Analytics are a new source of information that teams/athletes use to predict and better anticipate what the other team is going to do; they provide athletes the situational probability of an outcome in a given situation. These statistics go beyond the basic relationships taught in youth baseball (i.e., pitchers throw curveballs when ahead in the count) and instead give information on a specific pitch locations as it relates to variables such as pitch type, pitch count, handedness of the batter and game situation (Gray, 2015, p. 1377). Currently, baseball is one of the few sports to be analytically driven; they have increased their collective efforts to devote additional time and energy towards analytics to provide more detailed statistical information for teams to use. There are 27 out of 30 teams in the MLB right now that have at least one person with some type of role in analytics (Baumer, 2014). With the role of analytics expanding every year in the MLB since it was first introduced, organizations now have a tremendous amount of information at their fingertips; the only question now is how might they utilize it.

Major League Baseball teams receive bundles of data every day on opposing teams from companies like Statcast AI, real-time machine learning stats produced by Amazon Web Services. This information gets passed out to players in the organizations in a long spreadsheet that the players can choose to view online or on paper (Anderson, 2017). There has been a lot of push back from current Major League Baseball players on

the use of analytics due to the vast amount of information they are given as part of the game plan against opposing teams each night. The information is readily available to players, yet according to A.J. Ellis of the Los Angeles Dodgers, it is up to the players whether or not they actually use the information they receive for their game plan (Lemire, 2015). Furthermore, Chase Headley of the San Diego Padres, like many other baseball players, is appreciative of all the advanced information, but worries that the information can go beyond what is necessary and is over their heads (Lemire, 2015).

Overall, teams have found a great deal of success in building franchises around analytics; however, they have not found a way to present this useful information to their players in a way they can understand it and trust it in critical situations. The current study proposes to investigate whether providing various amounts and different types of analytic information to baseball players affects their ability to retain statistical information.

Background

Analytics provides athletes the situational probability of an outcome in any given situation. There have been a variety of studies conducted on the relationship between anticipation and baseball performance. Morris-Benelli and colleagues (2018) examined the effect of prediction of pitch type on batting performance. Their findings showed that anticipation of pitches helped increase performance in batting and resulted in fewer incorrect swings when given the correct prediction (Morris-Binelli, 2018). On the other hand, in Gray's (2017) study on the ironic and reinvestment effects in baseball pitching, he found contradicting results when pitchers were told where to throw the ball. In this study, 24 pitchers were given the "hot" and "cold" zones of simulated baseball pitchers. The results showed that when given information on where to throw in the zone

the pitchers who were given both hot and cold zones hit less targets than pitchers who were only shown the cold zones (Gray, 2017). Both Rob Gray and Khaya Morris-Binelli provide background information on the effects of how certain types of information can increase or decrease athlete performance.

Additionally, Gray (2015) conducted a study on the effect of the amount of information an athlete can handle before a given task. To do this he took 36 college baseball players and put them through a batting simulation that gave them the pitchers up to date statistics. There was a control group that received no analytic information, a build-up group that received up-to-date information after each at bat, and a full information group that received every statistical probability at once (Gray, 2015). Gray found that giving baseball players the full amount of information beforehand translated to an immediate performance advantage compared to the other conditions; however, over time the full information group began to forget the information and their performance suffered, whereas the build-up group was given direct information after every at bat that allowed them to re-evaluate their hitting strategy and retain the information throughout the study (Gray, 2015). Similarly, McPherson and MacMahon (2008) studied how well baseball players retain information while preparing for at bats. They did this by recording verbal reports of participants when asked to prepare to bat. In this study they found that "expert" batters were able to retain and remember more information than novices due to sport specific metacognitive strategies. These "expert" batters were able to encode information based on the current game situation, his team's current needs or goals, and his strengths or weaknesses as a batter which made retrieval and monitoring of current events easier for them compared to novices (McPherson & MacMahon, 2008). Both the

results found in Gray's "Moneyball" study and McPherson provide insight on the amount of information an athlete can handle before each task and how players retain the information provided to them.

Furthermore, Farrow and Reid conducted a similar study investigating how situational probability information affected the anticipatory responses of various aged and skilled tennis players. There were two groups used in this study, a younger less skilled group and an older more skilled group; each group was put through the same task of identifying serve location. A video-based tennis serve anticipation task was utilized, in which situational probability was artificially created. Participants were shown serves from the perspective of the receiver and tasked with predicting the location of tennis serves on a plasma touchscreen. Serves were randomized into sets and games; however the first serve in every game was presented in the same location. Farrow and Reid found that after the ninth game the older tennis players were significantly faster at anticipating the first serve of each game than the younger tennis players; the older tennis players were more aware of the relationship between serve location and game score, which eventually allowed them to better anticipate what was coming (Farrow & Reid, 2012). Thus, age and skill may influence an athlete's ability to anticipate what is coming.

On the other hand, there has also been various research on the effects different types of learning have on comprehension of information. Currently, baseball players receive mass amounts of analytic data in the form of spreadsheets from their organizations every couple of days for new opponents; however, this method of learning new information has been proven ineffective (Anderson, 2017). Richard Mayer is most well-known for his study on multimedia for e-learning; he investigated whether providing different forms of

presenting information (narration, animation, or a combination of both) to college students affected their ability to learn and comprehend the material. Based on task condition, participants were taught how a bicycle tire pump works followed by a problem-solving test that recorded comprehension of material. Comprehension was measured by the participants performance on the problem-solving test. Mayer found that people learn more deeply from words and pictures than from words alone (Mayer, 2017). Mayer refers to this important finding as the modality principle, which actually can be used to argue against current presentation methods of analytics in baseball today. On the other hand, Richard Mayer also found in his study with Ruth Clark that this modality principle does not always hold up. There are certain circumstances in which a subject may actually learn better from one source of learning rather than a combination of sources. Factors like experience and difficulty of information being presented could affect whether the modality principle is effective. For instance, the information being presented may only be useful to the learner in one format (i.e. text). In this case adding another form of information like graphics may confuse the learner or overload them cognitively, which can result in a decrease in retention and comprehension of material. He refers to this principle as the reverse modality principle, thus it could be the case that some athletes may learn better from one form analytics (i.e. only spreadsheet numbers) rather than a combination of the two (Clark & Mayer, 2011). Overall, Mayer's findings provide implications for my study as to what method of presentation results in the best ability to retain information.

Overview

Based on the above research, because gradually implementing small amounts of analytic data has been shown to be effective in simulated batting tasks (Gray, 2015) and providing information in the form of words and images to college students has been shown effective in classroom environments (Mayer, 2017) and simulated batting tasks and sport specific online based retention tests are similar in that they both study baseball players' ability to retain information and classroom settings and online based retention tests are similar in that they both conduct research on participants measuring comprehension of material through a computer test, it is hypothesized that providing analytics to baseball players gradually in small amounts with a combination of graphs and numbers would be useful in improving baseball players' ability to retain statistical information. Based on this, the current study predicts that when smaller amounts of analytic information in the form of a combination of words and images is presented to baseball players in an online based retention test it will positively impact the baseball player's ability to retain sport specific information, thus increasing their ability to anticipate what is coming. The aim of this study was to find a better way to present analytics to players in a way they can actually understand it and utilize it within a real game scenario. To do this, I created an online retention test for baseball players that focuses on retention and comprehension of statistical information as well as an online survey that records baseball players' feelings about different presentation methods of statistics. The study aimed to test the impact that various amounts and types of information provided to players has on their ability to retain statistical information.

CHAPTER 2

METHODOLOGY

Design

Experimental design of this online study followed a 2 by 2 within subjects factorial design. The amount of information and type of presentation method was varied based on condition; there were 4 total conditions possible in this study. Amount of information had 2 levels: High amount of analytic information or low amount of analytic information. High amounts of analytic information reflects the current methods of presenting analytics to players; it contains excess information like speed, spin rate, HR allowed, etc. Low amount of analytic information consists of only key information. This is only information that pertains to the task of batting in a game scenario; for example, information on pitch location, pitch type, and percentage of called strikes. The purpose of varying the amount of information is to observe how the amount of information affects retention of athletes. Furthermore, the type of presentation method had 2 levels: Numerical and numerical plus graphical. Numerical presentation method represents the current method of presenting analytics to players. Information in this condition was shown in a spreadsheet that only contains numbers. On the other hand, numerical plus graphical consists of information shown in a combination of spreadsheet numbers and a zone map to display pitch distributions of pitchers. The purpose of varying the type of presentation method is to observe how presentation of information affects athletes ability to retain statistical information.

Participants

A total of 34 randomly selected individuals participated in a 1 hour long online study. Total sample size was based on a power analysis conducted for the proposed 2 by 2 within subjects factorial design, which resulted in a sample size of 34 total participants. The participants were recruited from Arizona State University and the surrounding community. Participants were required to be fluent in English, be older than 18 years of age, and have some sort of baseball or softball experience. All participants who complete the study were automatically entered in a random prize drawing, which includes a $10 amazon gift card (5 gift cards will be awarded).

Materials

The online based study consists of a brief tutorial section in the beginning explaining task objectives and goals, 4 sections containing each a series of analytic information to review followed by a single comprehension question and a 10 question retention test, and a final survey at the end to record subjective preferences based on their experience during the study. Participants were asked to go through all 4 conditions in this study; conditions varied based on different amounts of information and different types of presentation methods. After completing an online consent form, participants in the study were given access to the study via Qualtrics.

Tutorial Section

This is the first section that the participant went through after completing the online consent form. The purpose of this tutorial section in Qualtrics was to explain task objectives, goals, performance scores, and provide examples of what to expect. In order to ensure that each participant receives the information in the same way, the tutorial section is the same text and images for each participant, thus every participant will see

and read the same information before starting the study. Additionally throughout the tutorial, check-up questions were put in to make sure the participant is paying attention. This helps ensure they adhere to instructions in online survey. Additionally, in an attempt to make sure participants listen and pay attention throughout the study they were asked in the tutorial section to take the survey in a quiet room without distractions and to remove any unnecessary material from the room. Furthermore, the actual tutorial section includes one image displaying how the analytic information will be provided to the participant (see Figure 1) and three example questions that the participant could see in the retention test (see Figure 2).

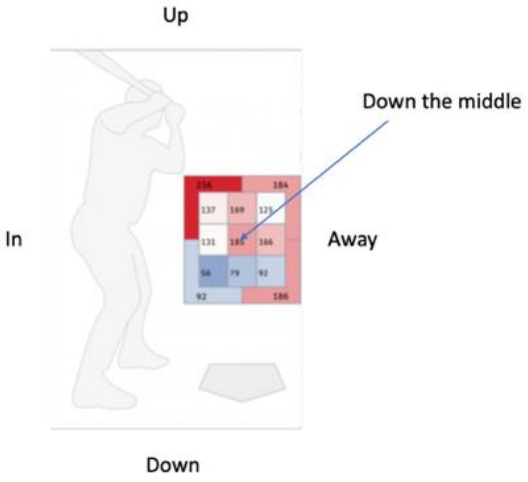

Figure 1. Tutorial Example Analytic Graph

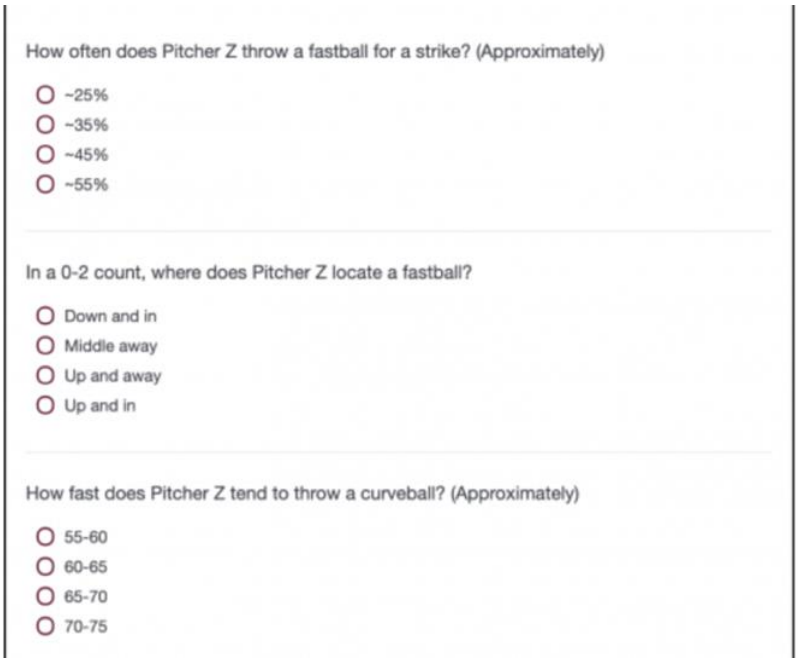

Figure 2. Tutorial Example Questions

Analytic Review Section

All statistical information and graphs came from baseballsavant.mlb.com (William, 2020). A total of four pitchers were randomly selected off of their website to use as the simulated pitchers in this study. Each pitcher in the database was given a Pitcher ID letter to keep them anonymous; statistical information on fastball and curveball statistics for all possible counts were taken for all 4 pitchers to be used in each condition. More specifically, review sections displayed statistical information on pitch specifics like called strikes and ball percentage, speed, spin rate, etc. (see Figure 3). At the beginning of each analytic review section, an image of the batter's view is displayed at the top. Analytics were given to the participant as if they were a right handed batter, which is explained in the tutorial section (see Figure 1). Therefore, left handed batters may be confused on where pitches are located; however, by utilizing the seductive detail principle and constantly placing an image that reminds the participant the viewpoint of

the analytics decreases the chance for confusion. Each participant takes the role of a potential batter and is tasked with memorizing as much information as possible before the analytics disappear (5 minutes). Based on condition, participants are shown various amounts and types of information on simulated pitchers. There were a total of 4 conditions that each participant ran through each condition; therefore, there were a total of 4 analytic review sections (1 for each condition). Before beginning the retention test for each section, participants were allotted 5 minutes to review and memorize a series of analytic information on a simulated pitcher. During the 5 minutes review section, the participants were asked 1 open-ended question measuring comprehension of material and transfer. The online retention test was reviewed by an expert in the field of baseball and other researchers to improve construct validity. After the 5 minute analytic review section was up, Qualtrics redirected the participants to the retention test portion of the study.

Pitcher D Fastball Distribution

Count	Pitch Location		Total # Pitches	# of Strikes	# of Balls	% of Strikes
(0-0)	Down and away		362	219	143	60.4972376
(0-1)	Up and in		163	88	75	53.9877301
(0-2)	Up and away		55	24	31	43.6363636
(1-0)	Down and away		161	102	59	63.3540373
(1-1)	Up and away		137	73	64	53.2846715
(1-2)	Up and away		84	54	30	64.2857143
(2-0)	Up and in		63	45	18	71.4285714
(2-1)	Up and in or Down and away		94	61	33	64.893617
(2-2)	Up and in		88	58	30	65.9090909
(3-0)	Middle in or Down and in		30	24	6	80
(3-1)	Down and away		56	41	15	73.2142857
(3-2)	Down the middle or Down and away		92	75	17	81.5217391

Figure 3. Analytic Information on Pitcher D

High Amount of Analytic Information Presented in Numeric. Participants in this condition were shown a high amount of analytic information in numerical form; this condition resembles the current method of presenting analytic information to players in pro sports. Information was displayed in this condition in a spreadsheet and provided statistics on a simulated pitcher's pitch distributions and tendencies. Specifically, the spreadsheet contains information about every possible count (12 possible counts) on two types of pitches (fastball & curveball) the batter could encounter, as well as pitch specifics (speed, spin rate, etc.) (see Figure 4). This condition, unlike other conditions, contains excess information in spreadsheet form.

Pitcher A Fastball Distribution

Count	Pitch Location	Total # Pitches	# of Strikes	# of Balls	% of Strikes	Speed (mph)	Spin Rate (rpms)	Batting Average Against (BAA)	Strikeouts (K)	Hits (H)	Home Runs (HR)
(0-0)	Down the middle	151	103	48	68.21192053	91	2,110	0.316	0	28	6
(0-1)	Up and In	85	46	39	54.11764706	93	2,180	0.263	0	18	0
(0-2)	Up and In	64	28	36	43.75	93	2,190	0.287	39	21	0
(1-0)	Down and In	199	143	56	71.85929648	90	2,100	0.304	0	47	9
(1-1)	Down	137	98	39	71.53284672	88	2,040	0.204	0	12	1
(1-2)	Down and away	66	37	29	56.06060606	93	2,200	0.231	18	14	0
(2-0)	Down the middle	182	98	84	53.84615385	90	2,100	0.406	0	45	3
(2-1)	Middle away	110	78	32	70.90909091	91	2,150	0.293	0	20	1
(2-2)	Down and away	160	58	102	36.25	92	2,180	0.258	68	9	0
(3-0)	Away	152	78	74	51.31578947	89	2,060	0.342	0	36	18
(3-1)	Down and away	72	48	24	66.66666667	91	2,120	0.274	0	18	2
(3-2)	Up	142	37	105	26.05633803	94	2,200	0.259	48	6	4

Figure 4. High Analytic Information Presented in Numeric Condition

Low Amount of Analytic Information Presented in Numeric. Participants in this group were shown a low amount of analytic information in numerical form. Similar to the first condition, the information displayed in this condition was still only numerical and came in spreadsheet format; however, it only provided key information. Specifically, this condition provided statistics on only simulated pitchers pitch distributions (see Figure 5). The spreadsheet will contain information about every possible count (12 possible counts) on two types of pitchers (fastball & curveball) the batter could encounter

from the pitcher. This condition provided less statistics than condition 1, thus it excluded information on pitch specifics.

	Pitcher B Fastball Distribution				
Count	Pitch Location	Total # Pitches	# of Strikes	# of Balls	% of Strikes
(0-0)	Down and away	131	79	52	60.30534351
(0-1)	Middle away	70	48	22	68.57142857
(0-2)	Down and away	67	25	42	37.31343284
(1-0)	Down the middle	65	47	18	72.30769231
(1-1)	Up and away	144	83	61	57.63888889
(1-2)	Up	91	18	73	19.78021978
(2-0)	Down the middle	78	58	20	74.35897436
(2-1)	Down and away	72	39	33	54.16666667
(2-2)	Down and in	34	23	11	67.64705882
(3-0)	Down	149	105	44	70.46979866
(3-1)	Up and in	109	72	37	66.05504587
(3-2)	Up and away	70	53	17	75.71428571

Figure 5. Low Analytic Information Presented in Numeric Condition

High Amount of Analytic Information Presented in Numeric and Graph. Participants in this group were shown a high amount of analytic information in combination of numerical and graphical form; this condition resembles the first conditions format in that it will display the same amount of information; however, it was provided in a combination of graphs and numbers. This condition utilized hitting zone maps to display pitch distribution as well as some spreadsheet information to display pitch frequencies and pitch specifics (see Figure 6). Zone maps display information on where the pitcher most frequently throws a pitch in a given count. It is hypothesized that by adding the visual to the spreadsheet of analytics, athletes will be able to better retain information rather than one source of learning alone.

Pitcher C Fastball Distribution

Count	Pitch Location		Total # Pitches	# of Strikes	# of Balls	% of Strikes	Speed (mph)	Spin Rate (rpms)	Batting Average Against (BAA)	Strikeouts (K)	Hits (H)	Home Runs (HR)
(0-0)	Up and In		504	334	170	66.2698413	97	2500	0.304	0	14	5
(0-1)	Up and In		188	134	54	71.2765957	95	2470	0.478	0	11	4
(0-2)	Up and In		90	60	30	66.6666667	95	2450	0.034	21	1	0
(1-0)	Down and In		154	98	56	63.6363636	95	2400	0.25	0	5	1
(1-1)	Down and away		132	96	36	72.7272727	94	2400	0.294	0	4	1
(1-2)	Up and away		119	75	44	63.0252101	96	2460	0.136	31	7	3
(2-0)	Down the midde		67	51	16	76.119403	95	2430	0.333	0	5	1
(2-1)	Up and In		101	77	24	76.2376238	96	2470	0.333	0	3	1
(2-2)	Up and In		184	143	41	77.7173913	95	2400	0.122	28	11	2
(3-0)	Down and away		35	23	12	65.7142857	94	2400	0.333	0	1	1
(3-1)	Up and In		71	52	19	73.2394366	96	2480	0.75	0	7	1
(3-2)	Up and In		115	94	21	81.7391304	97	2500	0.174	24	5	2

Figure 6. High Analytic Information Presented in Numeric/Graph Condition

Low Amount of Analytic Information Presented in Numeric and Graph. Participants in this group were shown a low amount of analytic information in numerical form. The information displayed in this condition was similar to the last condition (condition 3) in that it was shown as a combination of graphs and numbers; however, it only provided key information. That is, only information that the batter needs before facing the simulated pitcher; for example, pitch distributions for all possible pitch counts (12) for two types of pitches for the simulated pitcher (see Figure 7). Therefore, this condition differs from the other conditions in that it displays small amounts of analytic information in a combination of zone maps and spreadsheet numbers.

Pitcher D Fastball Distribution

Count	Pitch Location		Total # Pitches	# of Strikes	# of Balls	% of Strikes
(0-0)	Down and away		362	219	143	60.4972376
(0-1)	Up and in		163	88	75	53.9877301
(0-2)	Up and away		55	24	31	43.6363636
(1-0)	Down and away		161	102	59	63.3540373
(1-1)	Up and away		137	73	64	53.2846715
(1-2)	Up and away		84	54	30	64.2857143
(2-0)	Up and in		63	45	18	71.4285714
(2-1)	Up and in or Down and away		94	61	33	64.893617
(2-2)	Up and in		88	58	30	65.9090909
(3-0)	Middle in or Down and in		30	24	6	80
(3-1)	Down and away		56	41	15	73.2142857
(3-2)	Down the middle or Down and away		92	75	17	81.5217391

Figure 7. Low Analytic Information Presented in Numeric/Graph Condition

Retention Test

The retention test was constructed in Qualtrics and designed to measure participant's ability to retain analytic information. There were 10 total questions per retention test with a total of 4 retention tests for each condition. The condition order was randomized for each participant through Qualtrics; therefore, each participant received the conditions in different orders. The test itself has 4 different types of questions that can be asked based on condition: pitch type, location, call (strike or ball), and pitch specifics (speed, spin rate, etc.) (see Figure 8). Furthermore, the retention test measured the participants ability to retain analytic information on simulated pitchers. All questions in the test were multiple choice. Performance scores were based on the participants ability to answer questions on the retention test correctly; for every correct answer they received 1

point. In order to strengthen construct validity, I presented my retention test to experts in the field of baseball as well as other researchers who have designed tests in the past in order to ensure the questions asked were valid.

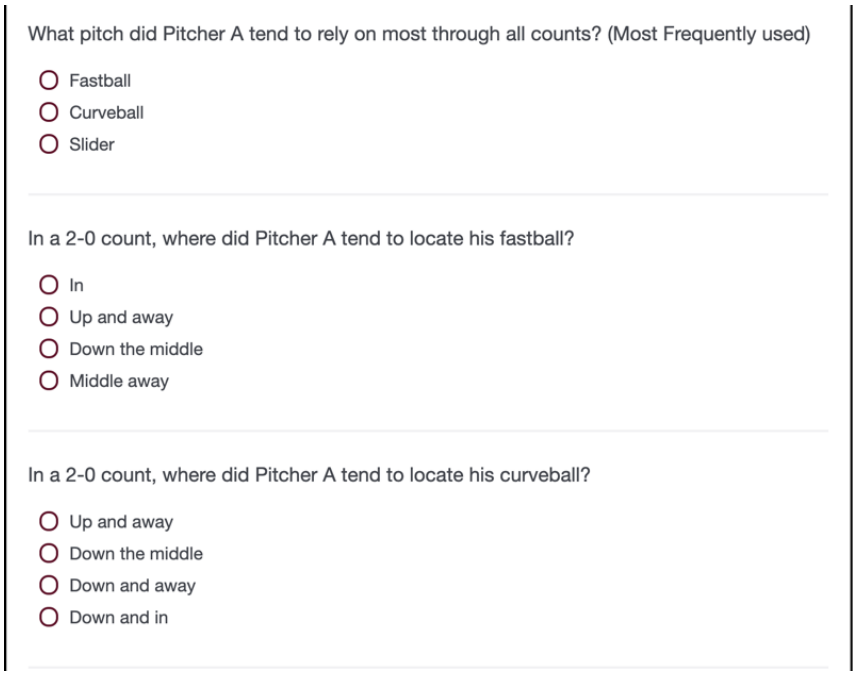

Figure 8. Retention Test

End Survey

The following survey was placed at the end of the study and asked for participants' subjective opinions and feelings about the study. There were a total of 5 multiple choice questions as well as one comment section if they felt the need to elaborate on anything (see Figure 9). The purpose of this end survey was to record participant preferences and feelings about the different types of presentation methods used throughout the study and understand their feelings towards analytics. Data points collected in the end survey were used to rank conditions based on preference and difficulty. Similarly to the retention test, the end survey presented to participants in the study is originally made; therefore, I

enlisted the help of other researchers who have created surveys in the past in order to improve construct validity and make sure the questions asked measured what I set out to measure.

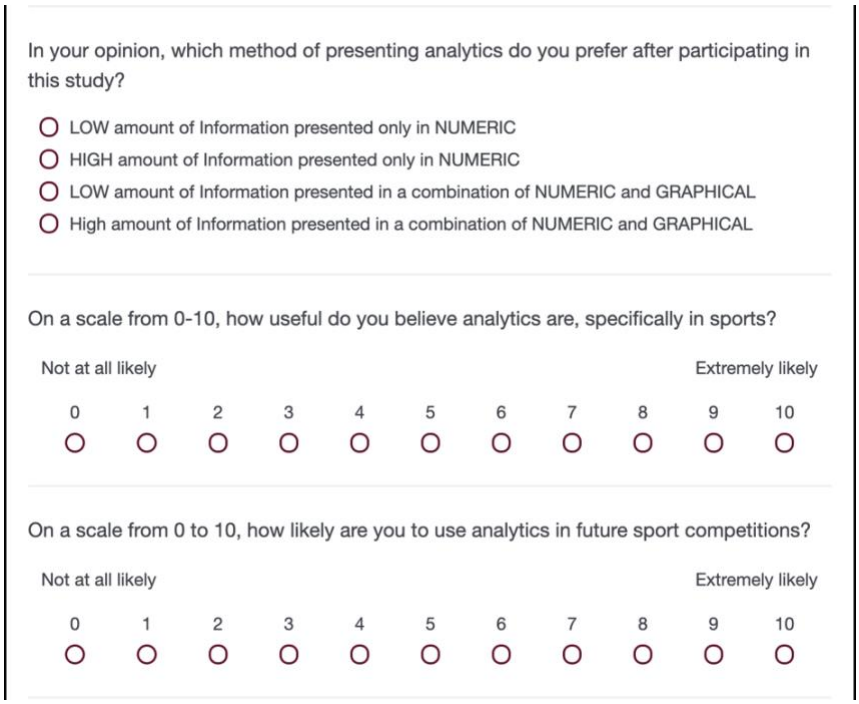

Figure 9. End Survey

Experimental Procedure

One participant at a time engaged in a 1 hour long study that included reviewing a 5 minute introduction tutorial, four 5 minute analytic review sessions at the beginning of each condition (4 total conditions), a single comprehension question measuring the participants understanding of the analytic material and their ability to apply the information to a real game scenario, followed by a 10 question retention tests (4 total, 1 for each condition), and a final 5 question subjective survey about their experience during the study. After obtaining informed consent, participants were given a link to the online study via Qualtrics. Participants then started the study by going through a short tutorial

on task objectives and description; each participant received the same tutorial. This tutorial informed the participants on the task narrative, how the information will be shown to them, how long they will have for each section, their goals, and performance scoring. Participants went through all 4 conditions; each condition was formatted the same, but the presentation method and amount of information provided varied. The order of the conditions were randomly assigned in Qualtrics for each participant.

Following the tutorial, participants partook in the retention based portion of the study. In this section of the study, participants received statistical information about simulated pitcher's tendencies; their job was to act as the batter who will face these pitchers at a later date. Based on condition, they must memorize the analytic information provided to them and answer questions specifically about the simulated pitcher's tendencies. Furthermore, the retention based section of the study will give participants 5 minutes to review the analytic information provided in the Qualtrics survey followed by a single comprehension question and a 10 question test measuring retention for each condition. After the conclusion of the retention tests, participants will be asked to complete a short 4 question subjective survey at the end asking for their preferences and how their experience went during this study.

CHAPTER 3

RESULTS

Retention Effects

Figure 10 shows the mean retention scores and standard errors for the different methods of presenting analytics of the experiment separated by experience level. Table 1 gives the means and standard deviations for each condition. Retention score data was normally distributed and did not have statistically significant sphericity or skewness.

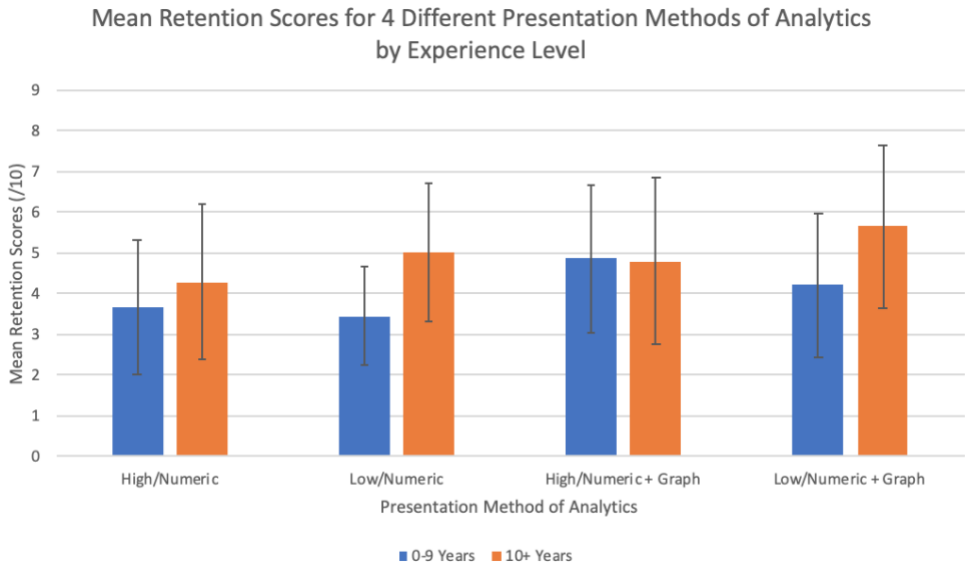

Figure 10: Mean Retention Scores for the Different Methods of Presenting Analytics by Level of Experience. Error bars are Standard Errors.

A 2x2x2 mixed ANOVA, with playing experience (0-9 years, n = 20, 10+ years, n = 14) as the between subjects variable and information type (numeric, numeric + graph) and information amount (low, high) as the within subjects variables, was used to assess whether different methods of presenting analytics produced differences in athletes

retention scores. Mauchly's test indicated that the sphericity assumption for the analysis was satisfied. The ANOVA analysis indicated statistically significant main effects of information type [$F(1,33) = 10.388$, $p = 0.002$, $\eta^2 = 0.245$] and experience [$F(1,32) = 4.842$, $p = 0.035$, $\eta^2 = 0.131$]. As shown in Figure 11, retention scores were significantly higher in the numeric + graph condition (by 0.8 points on average) with a large effect size. As shown in Figure 12, retention scores were also significantly higher for more experienced players (by 0.9 on average) with a medium effect size. On the other hand, information amount was not statistically significant with a low effect size ($p = 0.489$, $\eta^2 = 0.015$).

There was a significant interaction between experience and amount [$F(1,32) = 5.519$, $p = 0.025$, $\eta^2 = 0.147$]. As shown in Figure 13, this effect occurred because retention scores were significantly higher for more experienced baseball players in the low amount of information condition (by 1.5 points on average) with a large effect size [$t(1,66) = -3.638$, $p = 0.001$, $d = 0.88$]. On the other hand, there was no significant difference between the high and low experience groups when a high amount of information was given with a small effect size ($p = 0.539$, $d = 0.151$). The interaction between type of presentation method and experience ($p = 0.407$, $\eta^2 = 0.022$), type of presentation method and amount of information ($p = 0.792$, $p = 0.612$, $\eta^2 = 0.002$), and the 3 way interaction between type, amount, and experience ($p = 0.612$, $\eta^2 = 0.008$) were all not statistically significant and had small effect sizes.

Table 1

Means and Standard Deviations of Retention Score Data

By Experimental Condition

Condition	*Experience*	*M*	*SD*	*N*
0-9 Years' Experience				
High Numeric		3.65	1.66	20
Low Numeric		3.45	1.19	20
High Numeric plus Graph		4.85	1.81	20
Low Numeric plus Graph		4.2	1.76	20
10+ Years' Experience				
High Numeric		4.29	1.90	14
Low Numeric		5.0	1.71	14
High Numeric plus Graph		4.79	2.04	14
Low Numeric plus Graph		5.64	1.98	14

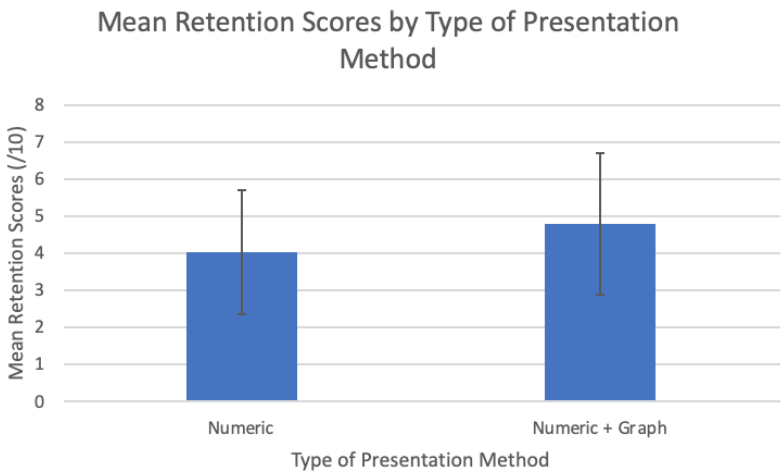

Figure 11: Mean Retention Scores Based on Type of Presentation Method Given. Error bars are Standard Errors.

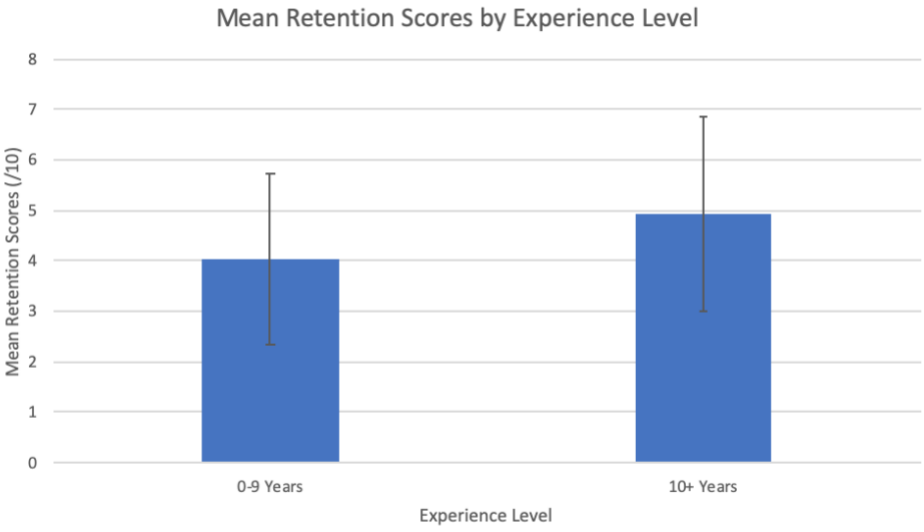

Figure 12: Mean Retention Scores Based on Baseball Experience Level in Years. Error bars are Standard Errors.

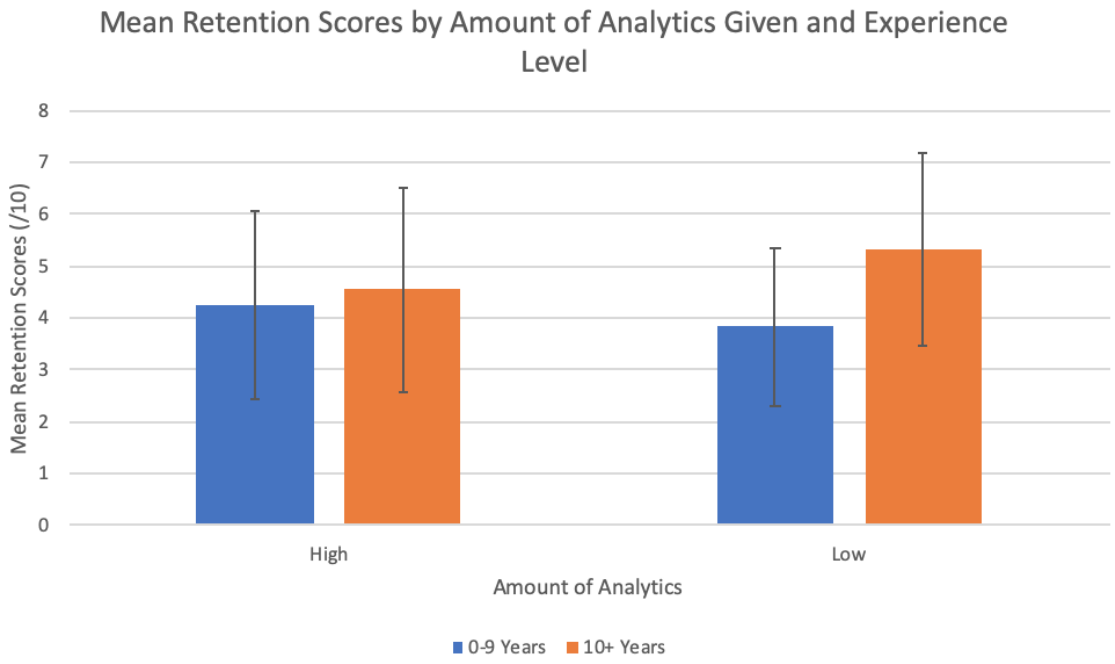

Figure 13: Mean Retention Scores Based on Amount of Information Given and Level of Baseball Experience. Error bars are Standard Errors.

CHAPTER 4

DISCUSSION

The aim of this study was to investigate whether different methods of presenting analytics to athletes produced differences in retention scores. It was thought that athletes that were presented less information in the form of numeric and graph would produce significantly better retention rates. The results of this study provided some evidence consistent with this hypothesis. For the condition that gave analytics in combined numeric and graphical form, athletes produced significantly higher retention scores than when only given numeric information. This finding is similar to Mayer's research on multimedia; people learn more deeply for a combination of words and pictures as opposed to when it is presented in only one modality (Mayer, 2017).

On the other hand, the results of the study did not show any significance for the amount of information alone. It was thought, based on previous research on situational probability use in baseball, that when athletes are given less information they would theoretically perform better than when given excess information (Gray, 2015). Unexpectedly, there was no significant difference in retention scores between low amount of information given and high amount of information given. A potential problem with presenting new information to learners is that the cognitive demands needed for learning a new task may exceed the capacity of the cognitive system (Mayer & Moreno, 2003). As discussed below, it is possible that this occurred because there's not enough information given in the "high" condition to cause an impairment in performance or there was too much information given in both conditions that caused impairment in performance for both conditions. This is also likely the reason why the predicted

interaction between the amount of information given and type of presentation was not significant.

The present study also provided evidence that playing experience influences how analytic information should be presented. On average, participants with more baseball experience performed better in every condition given to them compared to their less experienced counterparts. This can be attributed to the fact that athletes with more baseball experience were able to take on more baseball specific information than less experienced athletes. Presumably, athletes with more experience already had pre-existing knowledge of what statistics like batting average, strikeout rate, etc. meant so they could spend more of their time focusing on memorizing the actual analytics. This observed effect has been seen in other research as well; for instance, research on chess player's ability to recall different chess positions within a game. William Chase and Herbert Simon observed master chess players recall almost perfectly chess positions presented for only a few seconds. They found that master players use a chunking method by recognizing familiar patterns to help encode information in memory in chunks, thus allowing them to perform significantly well in recall tasks (Chase & Simon, 1973). Therefore, more experienced baseball players in this study may have been able to use this same chunking method to help them perform better in the retention tasks than less experienced players.

The results of the study also indicated a significant interaction between level of experience and amount of information given in terms of retention performance. More specifically, when more experienced athletes were given low amounts of analytic information they performed significantly better than less experienced athletes while there

was no significant difference between the experience levels in the high information condition. Notably, this occurred even though, as discussed above, the main effect of information amount was not significant. This finding was most likely due to the familiarity of the topic. The study was baseball specific and gave analytics only on the topic of baseball and as the results indicate, less experienced athletes performed about the same no matter the amount of information given; however, more experienced athletes performed significantly better in the low amount condition. It is hypothesized that this occurred because the experienced players were able to use their existing knowledge to "fill in the gaps" and derive more information in the low condition as compared to less experienced players. This effect may have occurred because of the prior knowledge principle; design principles that help low-knowledge learners may not effect high-knowledge learners in the same way. Kalyuga's research provides evidence that suggests presentation of information should be different for learners based on the level of acquired knowledge in a specific domain (Kalyuga, 2005).

Finally, athletes with more experience tended to give more detailed answers to the comprehension question. Athletes with more experience gave a more detailed approach where less experienced athletes tended to give short answers. This could indicate how athletes with different levels of experience attempt to memorize analytic information. Athletes with more experience gave more detailed approaches; therefore, they most likely look at the analytics given and memorize the information based on real game scenarios. As discussed above, master chess players had been observed using a chunking method to help encode information in memory in chunks, which allowed them to perform

significantly well in recall tasks (Chase & Simon, 1973). Therefore, the effect observed here could have also been the result of the chunking method for more experienced players

Practical Implications

The findings of the present study have important practical implications for how analytic information is used in baseball. Currently, baseball players are given an advanced stat sheet listing out numerous situational probabilities; some related to the task of batting and some just based on the pitcher's abilities and stats. The plethora of information given can be overwhelming and based on previous research can overload the athlete causing them to disregard analytics entirely (Lemire, 2015).

The fact the type of presentation method was found to be significant in previous research and this study provides strong evidence that having multiple modalities when presenting new information to someone seems to be more effective than one alone when looking at retention of material. Also interesting in this study was the significant effect of experience and the significant experience x amount of information interaction. As discussed above, more experienced athletes performed better over the entirety of the study and specifically when a low amount of information was given. This can hypothetically be attributed to the more experienced baseball player's familiarity with the information given compared to less experienced players. The ability for more experienced players to relate the information to real game situations faced previously could be seen to be very important when looking at their ability to retain baseball information compared to non-baseball players. These findings suggest that it might be most effective if analytic information were presented in different ways to players at different levels of baseball.

Limitations

A plausible limitation encountered during this study was the within subject design. Within subjects design was a good idea because of the small targeted population that I was trying to recruit, but I think since participants had to go through each condition they started to experience "fatigue" effects or gave up in the final condition because it was too taxing or they just wanted to finish. The study wasn't overly difficult or overwhelming, but I could see in the participants getting to the third or fourth retention test and just wanting the test to end or get tired of looking at all the analytic data. I think it is possible that some of the variability I saw within each condition was due to the fatigue effects. However, after conducting a fatigue analysis I found that fatigue was not present in this study. I would switch it to between subjects design if I could get the right number of participants. This would reduce any possible fatigue effects and allow each participant to only go through about a 20-minute study. Now this would introduce other problems that weren't relevant in a within subjects design, but I just think the variability was too high in each condition that was attributed to the amount of information of study.

The fact that the study was conducted online and was multiple choice also could have been a plausible limitation. It was hard to make sure that the participants were taking the study seriously, especially since it was administered online. I was unable to physically observe participants to make sure that they were fully engaged and completing all the task objectives. After the pandemic is over, it may be more beneficial to conduct this study in person rather than online. Additionally, I received a lot of good information through the comprehension question that was open ended, but I believe that the multiple-choice section measuring retention could have caused variability. I think the possible

variability had to do with the fact that the multiple choice questions didn't completely measure the participant's retention ability. Participants could just guess on 1 of 4 answers and have a 25% chance of getting it right. I was not expecting participants to get all 10 questions correct in each retention test, but I have a feeling that because they were multiple choice questions the participant at times selected the "closest answer" rather than answer based on their "true knowledge". Maybe making the question open ended rather than multiple choice may have helped in forcing the participant to think more critically and put their best answer rather than the one they think it is based on the process of elimination in order to effectively measure retention. The original test was reviewed by an expert in the field prior to administering it in the study and the data does not show any fluctuations or changes because the retention test was multiple choice. Therefore, this limitation may not have affected the data as first perceived.

Another plausible limitation to the study was that we did not access transfer. The study was focused on investigating the effects of multiple modalities and amount of information on retention rates; however, it is hard to generalize results to game performance without accessing how effective athletes could use the information given to them when actually hitting. The study lacked external validity; retention of analytic information tests provide evidence that athletes learn new information better in multiple modalities rather than just one and level of experience affects how well an athlete learns new information especially when given different amounts of information. However, the dependent measures of this study are not a good indicator of how athletes will perform in real game scenarios rather they indicate the best method to present analytic information to athletes. Instead of just measuring retention, I would also create a measure of hitting

performance that would measure transfer. If applied to the current study, athletes would partake in the retention portion of the study followed by a hitting performance test. The hitting performance test would measure how well athletes were able to apply what they learned in the retention portion to real game scenarios. Investigating whether transfer occurs on top of retention scores would increase external validity because the dependent variables would better reflect what baseball players would perform when doing the real thing; therefore, it would be easier to generalize the results of the study.

Conclusions

In sum, the goal of this study was to assess whether presenting analytic information to baseball players using different modalities and in different amounts would influence their ability to retain information. Findings providing some evidence to support my hypotheses. There was an observed main effect for type of presentation method, but there was no observed effect for the amount of information given or the interaction between type and amount. Results provided evidence that numeric plus graph conditions were more effective than just numeric. It can be seen that the type of presentation method results found in this study were similar to Mayer's research on multimedia. Both found that providing information in the combination of two modalities was better than just one modality alone. On the other hand, results in this study did not produce any significance for the amount of information given like predicting previously.

Furthermore, the findings on level experience were very interesting. Results provided significance for the main effect of experience and the interaction between experience and amount of information given. The significant difference can be attributed to the participant's familiarity with the material given and knowledge base. More

experienced baseball players were able to relate to the baseball specific information given better compared to their less experienced counterpart.

In future studies, I think it would be very interesting to test this same experimental design but look at performance measures. The study would still present analytic data in different modalities and in different amounts, but rather than testing whether they retain the information test how they perform the next day in a simulated baseball scenario. I think that it would be interesting to see if athletes change their approach when facing a pitcher after shown analytic information in multiple modalities and different amounts. I would predict that athletes would perform the best when given information the form of multiple modalities and when only presented key information. I would expect this condition to alter their batting approach around the analytics given. Whereas, conditions that give excess information would see the athletes ignoring the analytics and facing the pitcher based on reflexes.

REFERENCES

Anderson, R. J. (2017). The surprising places MLB teams get their information from in the post Moneyball era. Retrieved from https://www.cbssports.com/mlb/news/the-surprising-places-mlb-teams-get-their-information-from-in-the-post-moneyball-era/

Abernethy, B., Gill., D. P., Parks, S. L., & Packer, S. T. (2001). Expertise and the perception of kinematic and situational probability information. Perception, 30, 233-252.

Alain, C., Proteau, L. (1980). Decision making in sport. In: Nadeau, C. H., Halliwell, W.R., Newell, K.M., Roberts, G.C., editors. Psychology of motor behavior and sport. Champaign IL: Human Kinetics. pp. 465–477.

Baumer, B., & Zimbalist, A. (2014). The Sabermetric Revolution: Assessing the Growth of Analytics in Baseball. Philadelphia: University of Pennsylvania Press.

Chase, W. G. & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55–81. https://doi.org/10.1016/0010-0285(73)90004-2

Clark, R. C., & Mayer, R.E. (2011). Applying the modality principle: Present words as audio narration rather than on-screen text. In E-learning and the science of instruction (pp. 115-121). San Francisco, CA: Wiley.

Farrow, D., & Reid, M. (2012). The contribution of situational probability information to anticipatory skill. Journal of Science and Medicine in Sport, 15, 368-373.

Gray, R., Orn, A., & Woodman, T. (2017). Ironic and reinvestment effects in baseball pitching: How information about an opponent can influence performance under pressure. *Journal of Sport & Exercise Psychology, 39*(1), 3-12.

Gray, R. (2015). The Moneyball problem: What is the best way to present situational statistics to an athlete? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *59*(1), 1377–1381. https://doi.org/10.1177/1541931215591230

Kalyuga, S. (2005). Prior knowledge principle in multimedia learning. *The Cambridge handbook of multimedia learning*, 325-337.

Lemire, J. (2015). Hey, data data: MLB teams face challenge delivering info to players. Retrieved from https://www.usatoday.com/story/sports/mlb/2015/08/26/mlb-analytics-data-reports/32410841/

Lewis, M. (2004). *Moneyball: The art of winning an unfair game*. London: W. W. Norton & Company.

Major League Baseball. (2019). *Statcast*. [online] Retrieved from http://m.mlb.com/glossary/statcast

Mayer, R. E. (2017). Using multimedia for e-learning. Journal of Computer Assisted Learning, 33(5), 403-423. Retrieved from https://doi.org/10.1111/jcal.12197

Mayer, M. & Moreno, R. (2003). Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educational Psychologist*, *38*(1), 43–52. https://doi.org/10.1207/s15326985ep3801_6

McPherson, S. L., & MacMahon, C. (2008). How baseball players prepare to bat: Tactical knowledge as a mediator of expert performance in baseball. *Journal of Sport & Exercise Psychology*, 30, 755-778.

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and instruction*, *14*(1), 1-43.

Morris-Binelli, K., Müller, S., & Fadde, P. (2018). Use of pitcher game footage to measure visual anticipation and its relationship to baseball batting statistics. *Journal of Motor Learning and Development, 6*(2), 197-208.

Willman, D. (2020). Baseball savant: Trending MLB players, Statcast, and visualizations. Retrieved from https://baseballsavant.mlb.com

Woodman, T., Barlow, M., & Gorgulu, R. (2015). Don't miss, don't miss, d'oh! Performance when anxious suffers specifically where least desired. The Sport Psychologist, 29, 213–223. doi:10.1123/tsp.2014- 0114 34

APPENDIX A

IRB EXEMPTION FOR HUMAN TESTING

Listed below is the approval/exemption form for human subject testing. The IRB had granted approval to conduct human subjects testing for this study.

EXEMPTION GRANTED

Robert Gray
IAFSE-PS: Human Systems Engineering (HSE)
480/727-1419
robgray@asu.edu

Dear Robert Gray:

On 5/12/2020 the ASU IRB reviewed the following protocol:

Type of Review:	Initial Study
Title:	Analytics in Baseball: Retention of Sport Specific Analytic Information Based on Various Presentation Methods
Investigator:	Robert Gray
IRB ID:	STUDY00011942
Funding:	None
Grant Title:	None
Grant ID:	None
Documents Reviewed:	• Analytic_Survey (Qualtrics).pdf, Category: Measures (Survey questions/Interview questions /interview guides/focus group questions); • Consent Form.pdf, Category: Consent Form; • Consent_Demographics (Qualtrics).pdf, Category: Measures (Survey questions/Interview questions /interview guides/focus group questions); • IRB Social Behavioral 2020 Gin 2.docx, Category: IRB Protocol; • Recruitment Statement.pdf, Category: Recruitment Materials;

The IRB determined that the protocol is considered exempt pursuant to Federal Regulations 45CFR46 on 5/12/2020.

In conducting this protocol you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).