Russell E Dean. A Study of Mobile Self-Tracking Applications. A Master's Paper for the M.S. in I.S. degree. April 2012. 36 pages. Advisor: Arcot Rajasekar

A growing number of individuals are collecting, organizing, analyzing, and sharing information about their personal behaviors and experiences. Numerous tools exist to assist this data collection, many of them applications designed for mobile devices. To explore the design details of these applications, 100 iPhone applications for self-tracking were examined and measured by a set of eight criteria. These criteria were then compared to user ratings, to determine their relative significance on user reaction to the applications.

The results of this model were mixed - seven of the eight criteria met hypothesized expectations, yet the overall predictive model for user ratings did not perform well. This paper discusses some of these findings as well as some of the possible problems with the model, and explores areas for future research.

Headings:

Mobile Applications

Personal Data Mining

Personal Informatics

Self-Tracking

Self-Surveillance

A STUDY OF MOBILE SELF-TRACKING APPLICATIONS

by
Russell E. Dean

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Science.

Chapel Hill, North Carolina

April 2012

Approved by

Arcot Rajasekar

1. Introduction

> “Of what lasting benefit has been man's use of science and of the new instruments which his research brought into existence?” (Vannevar Bush, Atlantic Monthly, 1945)

> “After filling out 50 happiness surveys, the program spits out a report…I’m at a slight loss for what to do with these results. Does this mean I should spend more time in bars and less time at work to optimize my happiness? …When I attend my first Quantified-Self Meetup, I discover that other members share my excitement and perplexity: They don’t know what to do with their data.”
> (Kashmir Hill, Forbes, 2011)

In 1945, Vannevar Bush wrote *As We May Think*, an essay often cited as an important precursor to the Internet and information technologies of the modern era. In this piece, Bush spoke of his concern with the “growing mountain of research” and for investigators of information “staggered by the findings and conclusions of thousands of other workers – conclusions which he cannot find time to grasp, much less remember…. (Bush, 1945)”. This complaint was not new to 1945 – at the time of Gutenberg’s invention of the printing press, scholars already complained of the amount of information they were required to sift through (Blair, 2010). What was new about Bush’s essay was the outlining of technologies to help information-seekers gather and organize relevant data, summarized in the design for the Memex machine, seen as a forerunner of the personal computer. In respect to the information-seeker, Bush saw this technology as an “enlarged intimate supplement to his memory” made necessary by the pace of information growth (Bush, 1945).

The expansion of data since Bush’s article has continued exponentially (IDC, 2007). In the last fifteen years, with decreases in the cost of digital storage and increases in processing power, companies and organizations have focused their efforts towards gleaning information from their expanding data warehouses – the set of processes used to discover new patterns from these large data sets is commonly known as “data mining” (“Data Mining”, 2011). At its core, the goal of data mining is to create new predictive data from the patterns found in existing data, which can then be used for more effective decision-making (Fayyad, Piatetsky-Shapiro & Smyth 1996).

As the demand for organization of information for businesses has grown, so too has the demand for more effective methods of personal information management (Bederson, Teevan & Jones, 2006). A growing number of individuals are collecting, organizing, analyzing and sharing information about their personal behaviors and experiences. There are many different terms used to describe this behavior in the corresponding research that seeking to analyze it – some include “self-surveillance”, “self-tracking”, “self-regulation”, “life-logging”, “personal informatics” or “personal data mining”. For the purposes of this paper, I use the term “self-tracking”, as this very clearly captures the act of gathering data about oneself.

An exemplar of the self-tracking movement is Daytum.com founder Nicolas Felton, who publishes Annual Reports each year with detailed statistical analysis of his daily life. Felton reports everything from amount of time spent on the subway, to the most popular topics of his conversations held throughout his year. Felton analyzes data from a variety of different devices or services, many of which automatically collect data for him, and boils everything down to a visualization-heavy report for public consumption. While many individuals may not consider themselves the same kind of self-tracker in comparison, a 2010 PewResearchCenter national survey on people’s interactions with health information found that 27% of internet users report having tracked information on their weight, diet, exercise routine, or other health indicators online (Pew Research, 2011). Certainly this data suggests that the self-tracking movement touches on a significant sector of the population.

Technology has arisen to meet the needs of self-tracking, and much of the research in the field focuses on the efficacy of technical devices for data collection (Sellen & Whitaker, 2010). A September 2006 patent application from Microsoft speaks of a "'personal data mining' system that would analyze information and make recommendations with the goal of aiding a person's decisions and improving quality of life." (Rowan, 2011) This technology would not be groundbreaking, however, as many such devices and services are currently available. Specially designed tools, such as the FitBit, are used to track sleep and movement levels automatically from the wearer. Websites such as Daytum.com or Flowingdata.com are devoted to the organization and transmission of personal data. The PewResearchCenter survey found that 9% of cell-phone users employed mobile health apps to track or manage their health as of 2010 (PewResearchCenter, 2011). While the prevalence of self-tracking technology and relevant research grows, Yau & Schneider argue that the research surrounding the field of self-tracking largely ignores the relative utility of tools or techniques (Yau & Schneider, 2009).

For the purposes of this study, I have chosen to survey a portion of the current self-tracking mobile application market, specifically apps available on the iOS platform, through Apple iTunes. Rather than examining the abilities of each application to collect data, I sought to find a model of common characteristics across self-tracking applications, and to determine the impact of these characteristics on the relative appeal of self-tracking applications through user rankings and popularity measures. Specific attention was paid to the ability for users to analyze the data that they collect, whether in the app itself, or through exporting of data for use with other tools. While I expected to find common threads in tracking apps, I sought to find how these common elements are truly received by the app-buying population.

2. Review of the Literature

In my review of the literature for this project, I first sought to understand the roots, motivations, and benefits of self-tracking. I then hoped to gain an understanding of the current state of self-tracking and modern technology. From this study, I hoped to glean a useful set of measures to use for study of current mobile applications for self-tracking.

Roots of Self-Tracking in Research:

In a 2011 piece, Wired UK editor David Rowan discussed his view of the state of self-tracking:

> "In our post-privacy world of pervasive social-media sharing, GPS tracking, cellphone-tower triangulation, wireless sensor monitoring, browser-cookie targeting, face-detecting, consumer-intention profiling, and endless other means by which our personal presence is logged in databases far beyond our reach, citizens are largely failing to benefit from the power of all this data to help them make smarter decisions…" (Rowan, 2011).

Rowan's piece mirrors much of the research on self-tracking, with heavy focus on evolving technology. While many of the data sources and methods he describes are realities of the last decade, research into collection and analysis of data about oneself has been around for much longer.

In 1977, Don Zimmerman and D. Lawrence Wieder wrote an early piece on the "Diary-Interview Method". This research method involved direct interviewing of subjects in conjunction with use of "annotated chronological record or log(s)" (Zimmerman & Wieder, 1977). This method differed from traditional research, in that participants were instructed to personally record data about their activities and experiences over a period of time, rather than having primary data collected from external sources. The method was favored because of the ability to minimize observer effects on data collection while at the same time improving subject recall – a subject was able to move "through his or her normal activities 'as if' the observer was not present, which is to say 'naturally'". (Zimmerman & Wieder, 1977) As the researchers instructed subjects on how and what to record, observers could still shape the data collection. Zimmerman and Wieder acknowledged that the study of "intimate journals", or unsolicited, naturally recorded documents from subjects, had been a useful element of psychological and sociological research dating further back.

One early advocate for use of these uncommissioned personal records was Gordon Allport, former president of the American Psychological Association. In his 1942 piece, "The Use of Personal Documents in Psychological Science", Allport speaks of the desire to "evaluate the increasing flood of personal documents", fully 70 years before David Rowan laments the same gnawing desire. Allport succinctly describes the value of self-tracking data:

> "As a self-revealing record of experience and conduct…usually, though not always, produced spontaneously, recorded by the subject himself, and intended only for confidential use…sometimes they represent distillations of the most profound and significant experiences of human life." (Allport, 1942)

While there is a long history of perceived value of self-tracking from the researcher's perspective – more recent studies have established the value of self-tracking to subjects themselves.

One common area of research on the effect of self-tracking is in the realm of weight-loss intervention. A 2011 study of 228 children with obesity found a significant increase in weight loss over a six-month period amongst children who regularly self-monitored eating and exercise, over children who did not. This study builds on a wealth of previous research, with similar studies establishing correlations between consistency of self-monitoring and long-term weight loss. This body of research has led some researchers to describe self-monitoring as "the cornerstone and the most effective technique used to help people lose weight." (Germann, Kirschenbaum & Rich, 2006).

Other studies have looked at the role of self-tracking in asthma, alcoholism, depression, and self-management of genetic disorders. (Giarelli, Bernhardt, & Pyeritz, 2009). In studies on self-tracking of experiences with health-care providers, unsolicited diaries were found to "provide insight into the experience of individuals interacting with the health care system" (Jacelon & Imperio, 2005). Wilde and Gravin, in their concept analysis of self-tracking find that overall "self-monitoring has been shown to enhance health behavior and lead to measurable changes in health-related outcomes" (Wilde & Garvin, 2007). Burke et al. summarize the possible reason for the value of self-tracking - "self-regulation theory posits that self-monitoring precedes self-evaluation of progress made towards one's goal and self-reinforcement for the progress made". Burke, Wang, Sevick, 2011). The review of self-tracking research yields a clear picture - not only are self-tracking records a useful tool for data collection for researchers, but also the act of self-tracking can have a direct positive impact on the subjects themselves.

Self-Tracking and Technology:

Turning attention towards current practices of self-tracking, a growing portion of relevant literature examines technological aspects of the field. Yau & Schneider (2009) argue that self-surveillance data collection has been made easier by advances in technology. Early examples of technologies used for self-tracking include peak-flow meters for asthma and glucometers for diabetes. (Wilde & Garvin, 2007). More recently, advances may take the form of increased access to data such as credit card bills or GPS information stored in mobile devices. Advances also result from tools specifically designed for self-surveillance data tracking, such as the biometric devices that automatically track sleep and movement levels of the wearer (Yau & Schneider, 2009). A survey of the literature suggests that these technological advances have had mixed results.

One recent study of self-tracking technology involved evaluation of personal electronic devices in use for self-monitoring in a pediatric weight management program (Cushing, 2011). Cushing et. al noted a dearth of research on these tools in respect to weight loss. While the sample size of this study left something to be desired, evidence suggested that PED use had a dramatic impact on the level of monitoring maintained by subjects. However, in a contrasting study of self-monitoring within a sample of obese adolescents, Kirschenbaum et. al found that only 50% of subjects continued self-tracking after one month, and less than 25% continued after six months (Germann, Kirschenbaum, & Rich, 2006). Others have suggested that the technological state of self-tracking leaves something to be desired. In their critique of "Lifelogging", Sellen and Whitaker suggest that while some of the implied goals of self-tracking may be met to a degree, the findings of much of the research into self-tracking technologies "imply that archival data may be less valuable than the considerable effort expended on these systems would justify (Sellen & Whittaker, 2010)." Again, the authors argue that analysis and utility of collected self-tracking data has yet to be fully studied.

Mobile Applications for Self-Tracking:

A search for studies of technological details of self-tracking shows many current studies in the works – for instance, an ongoing study on a mobile phone self-tracking tool for emotional health in adolescents, or a study of mobile applications for the self-tracking of inflammatory bowel disease.

While these studies center on the evaluation of individual self-tracking applications and their controlled test results, research into what makes for 'good' self-tracking applications for mobile devices remains lacking. Even so, hundreds of these applications are currently available for public use. This study would like to explore what the common characteristics of these mobile self-tracking applications are on the market, and what evidence there is that these characteristics have a positive impact on self-tracking.

3. **Research Design and Methodology**

Sampling - The desire to gain perspective on the entire market for mobile applications was tempered by the desire to keep certain elements external to the interest of this study constant. A decision was made to focus on free mobile iPhone applications for self-tracking. Staying within these bounds controlled for potentially confounding effects on user attitudes towards apps, such as differences in price, mobile operating systems, or specific mobile device quality.

With approximately 500,000 mobile applications currently available for sale through the “iPhone App Store” and a constant influx of new products, the number of self-tracking applications on the market is unclear. Apple, the administrator of the iPhone App Store, does not provide a specific category for self-tracking applications, and thus the relevant applications had to be found in the existing category structure. The top 120 mobile application lists of the 22 existing search categories in the iPhone App Store were scanned for applications indicating some level of self-tracking. These lists are ranked upon number of downloads as well as number of weekly and monthly users, and were chosen as a theoretically valuable organizational tool to find applications with a large enough number of reviews for study.

Mobile application titles and short mobile application descriptions were scanned for mentions of possible self-tracking terms and their derivatives, such as “tracker”, “counter”, “diary”, “log”, “monitor”, or “diary”. The vast majority of self-tracking apps were found in the “Health & Fitness” and “Medical” search categories of the iPhone App Store, while a few apps were found listed in other categories, such as “Lifestyle” or “Finance”. An initial sample of 100 applications was derived from this search.

It was quickly apparent that clear criteria were needed for determining if an application qualified as a “self-tracking application” for the purposes of this study. The working definition used was for “applications designed primarily for gathering and storage of data about a user, or about a user’s interaction with their environment.” Of the initial 100 applications in the sample, 19 were determined to not qualify as self-tracking applications by this definition. An example of applications that do not meet the criteria of the definition is pregnancy applications which ask for the expected due date of the child, and return relevant information for the given week of the pregnancy, but do not provide any means for tracking further information about the user (not primarily data-tracker). "Calculator" applications also were excluded from the final sample, such as Body Mass Index calculators, which ask for height and weight data, but simply use the information for one-time calculations rather than storage of information about the user.

While self-tracking applications were found to focus on a wide array of information, a few key categories appeared most often. The most common subjects of apps were weight control and dieting applications, exercise logging applications, and menstruation and fertility trackers. The iTunes App Store was checked further for self-tracking applications, which may not have been in the top 120 lists, but still included some of the previously mentioned vocabulary, bringing the total sample back to 100.

Criteria of Interest for Mobile Applications - From the literature review for self-tracking, a list was derived of attributes that might be found to have a significant impact on user reviews of the application ("criteria of interest").

Yau and Schneider suggest that visualization is the key analytical tool used in mobile applications – the presence of **visualization** is one criterion for study. From Sellen and Whitaker's suggestion that the analysis and utility of self-tracking data has not been fully studied, two more criteria are used to gauge the attention each mobile application pays to data analysis – whether **analysis** beyond visualization is present, and whether data collected by a mobile application can be **exported** for analysis outside of the application for external analysis. Zimmerman's emphasis that personal records are intended for confidential use raises the question whether **security** concerns may impact self-tracking applications. Burke's discussion of the importance of consistent and timely self-monitoring towards a target behavior suggest two more important criteria – the presence of **reminders** or notifications to help maintain consistent monitoring, and the ability to set **goals** as a target behavior for the self-tracking. Germann et. al mentions the significant impact of parental involvement in weight-loss tracking, suggesting that **sharing** of tracking information may also have an important impact on self-tracking. (Germann, Kirschenbaum & Rich, 2006). A final criteria of interest was found in reading through mobile developer guidelines for iOS developers – the strong suggestion that user-interface elements that hinder users' ability to use mobile applications should be avoided, such as screens requiring **early registration** from the user.

While these criteria of interest do not exhaust the possible significant elements of self-tracking mobile application development, they were expected to provide a broad look at important characteristics of the applications. Each application was downloaded, examined, and coded for the presence or absence of the eight attributes. The coding relied on the following questions for the eight criteria of interest –

- Early Registration: Does the application require registration of some kind upon initial use?
- Goal Setting: Does the application allow users to create any type of behavioral goal?
- Visualization: Does the application provide some level of visualization of data for the user?

- Analysis: Does the application provide any level of data analysis beyond visualization, or does it simply store the data?
- Export: Does the application provide a means for exporting of data for use outside the app, including providing web-access to data?
- Sharing: Does the application provide ways for users to share elements of the data that they chose, either through social media outlets or in-app discussion boards?
- Notifications/Reminders: Does the application allow users to establish reminders for keeping up with regular data tracking?
- Security: Does the application provide some level of security for the data that is being collected, in the form of password protection?

This study hypothesized that the presence of these criteria will have a positive impact on overall rating of the application, except in the case of early registration, which is expected to have a negative impact on overall rating. These eight criteria were culled and refined from a larger list of possible criteria, after test screening of a handful of applications. Additional criteria such as the ability to customize visualization of data were discarded, as very little evidence was found for this in the test applications. Measurement of app start-up time on use was also considered but removed from the final list, both because of the time-intensive nature of this analysis, and because of expected differences across mobile devices, networks, or even times of day. A choice to disregard the presence of advertisements was also made, as this was considered to be more of a logistical business consideration for mobile application developers, than a design decision.

Overall User Response to Mobile Applications: Along with the data on the above attributes collected for the web, a measure was needed for gauging user experience with the self-tracking applications. The ideal measure of mobile application raw sales and use data was not readily available for this study. Data on mobile application rankings were available for applications reaching the top 200, but there were a few compelling reasons to not focus on this data. The underlying algorithm behind the ranking process has not been made public, and the contribution of the number of downloads versus extent of usage is thus unclear. Relative ranking can also be skewed by the timing of popularity of other mobile applications on the market, thus a ranking may be artificially low because of an increase in competition in the market at a given time. Finally, longevity on ranking lists or top spots achieved on ranking lists can miss the details of the extent of popularity, as some apps may be consistently used at a moderate rate over time, while other apps may have high peaks of usage and trail off more quickly.

Publicly available user rankings of mobile applications were chosen as a better alternative. Each ranking is a simplified judgment of one user's feelings on the overall quality of the mobile application – this judgment is largely independent of relative rankings of other applications, and is made similarly on applications of varying popularity. While it is clear that rankings could be impacted by attributes external to the eight listed above, it was a goal of this study to determine if the eight proposed attributes proved to have a significant impact on the user experience with the applications – thus making these attributes worth design consideration for self-tracking mobile application developers.

Application ratings data was gathered from the appFigures online reporting platform, designed for reporting ratings information to mobile application developers. The ratings information was a snapshot collected during a ten day period in March of 2012. The number of stars given from 1 to 5 stars was recorded, and an aggregate overall rating computed.

4. Findings

Figure 1 shows the breakdown of categories of tracking focus for the mobile applications examined for this study. Exercise and weight maintenance related mobile applications accounted for 55 of the 100 apps, while 20 menstruation and pregnancy related apps were examined. The wide range of additional areas of tracking focus included allergies, blood pressure, cancer, diabetes, dreams, financial activity, medication, mood, parenting behaviors, sleep, stress, time management, and water consumption. In addition, one application was designed as an ‘all in one’ tracking application, allowing for customizable tracking of almost all of the areas of focus from the other apps. While self-tracking applications may theoretically be used for any data collection a user may be interested in observing, this survey of applications on the market shows that the vast majority of self-tracking apps fall into exercise/weight or menstruation/pregnancy fields.

Figure 2 shows the overall presence or absence of the eight previously mentioned criteria. Of particular interest was the gap between visualization and further data analysis. The use of some level of data visualization was present in 80 of the 100 mobile applications studied, by far the criteria most observed in the test sample. Analysis beyond visualization was present in only 24 of the 100 mobile applications – the least

Figure 1 – Frequency of Tracking Focus

all types	1
allergies	1
blood pressure	2
cancer	1
diabetes	5
dream	1
exercise	32
financial	2
medication	1
menstruation	13
mood	2
parenting	4
pregnancy	7
sleep	2
stress	1
time management	1
water intake	1
weight	23
Total	100

frequently observed of the test criteria. Clearly, visualization is the preferred method of data analysis provided in the sample of free self-tracking mobile applications, while further data analysis is somewhat rare.

The remaining six criteria were observed at the following levels in the 100 apps in the sample: Export (51), Security (48), Goals (43), Sharing (43), Reminders (38), and Early Registration (28). While there were many different combinations of these elements across each of the applications, there were some general trends seen within focal areas of tracking. Figure 3 shows the presence or absence of the eight criteria in only weight and exercise related mobile applications, as well as the differences in percentage from the same calculations done on the entire sample of applications. Several significant

Figure 2 – Aggregate Percentages for Criteria of Interest

Criteria	Presence in Applications
Early Registration	28%
Goal Setting	43%
Visualization	80%
Analysis	24%
Export	51%
Sharing	43%
Reminders	38%
Security	48%

differences exist – goal setting is present in 64% of weight and exercise applications, while present in only 43% of all self-tracking applications in the sample, analysis beyond visualization is present in only 9% of weight and exercise applications versus 24% in the entire sample. The sample sizes of the various focal areas were too small for further analysis in this study, but these differences would be of interest in future research.

Figure 4 compares the mean rating average for applications meeting a specific criteria against applications not meeting that criteria. In almost all of the categories, the mean rating moves in the hypothesized direction (negatively for early registration, and positively for the other criteria). The only exception to this concerns the goals rating, which actually showed a decrease in rating in apps with goal setting features. With most

Figure 3 – Aggregate Percentages for Criteria of Interest (weight and exercise)

Criteria	Presence in Applications	Difference from entire sample %
Early Registration	42%	+14
Goal Setting	64%	+21
Visualization	84%	+4
Analysis	9%	-15
Export	51%	0
Sharing	55%	+12
Reminders	24%	-14
Security	42%	-6

criteria, the difference in mean rating average between presence and absence of a criteria is less than two-tenths of a rating pt, the exceptions being reminders (~.48 increase), visualization (~.39 increase), and analysis (~.26 increase).

These increases suggest that there is a measurable difference in user response to self-tracking mobile applications that meet specific studied criteria. Regression analysis

Figure 4 – Mean Rating Average (5pt scale) by Presence and Absence of Criteria

Criteria	Average Rating when Present	Average Rating when Absent
Early Registration*	3.673	3.706
Goal Setting	3.644	3.737
Visualization	3.776	3.382
Analysis	3.897	3.634
Export	3.731	3.662
Sharing	3.806	3.615
Reminders	3.994	3.515
Security	3.801	3.601

*Hypothesized negative result when present was used to attempt to capture the independent effects of each of the variables on the overall rating average of these applications. Regression analysis is a statistical method for analyzing the impact of several variables on a dependant variable. The power of regression analysis is the ability to isolate the effects of each variable, with other variables remaining constant. Using reminders as an example for translating regression results – the .375 coefficient means that all other variables being equal, the presence of reminders results in a .375 increase in rating average. Significance level of the coefficient is also calculated, to establish how certain the model is that the relationship between the dependent and independent variable is not random. A significance level of less than .05 is desired, or a 95% confidence in the results

Figure 5 shows the results of a linear regression model, using all eight of the criteria of interest to predict the rating average of a given mobile application. Unfortunately, this predictive model did not yield statistically significant results in the regression test, with an adjusted R-squared of .92, indicating statistical significance only at a 90% confidence level. Six of the eight variables showed coefficients consistent with the hypothesized direction, ex. when security is present, rating average goes up .019. However, only one of the eight variables was shown to have a statistically significant impact on rating average - the presence of notifications or reminders to help users keep up with regular data tracking. To check the direct relationship of reminders and rating average, the Pearson correlation coefficient was calculated between the two variables. The Pearson correlation coefficient is a measure of how closely dependent two variables

Figure 5 – Regression Analysis of original model

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.407[a]	.166	.092	.634617898157565

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.323	.162		20.553	.000
	Early Registration	-.085	.164	-.058	-.517	.606
	Goal Setting	-.024	.138	-.018	-.171	.864
	Visualization	.185	.190	.111	.971	.334
	Analysis	.175	.169	.113	1.036	.303
	Export	-.025	.135	-.019	-.188	.852
	Sharing	.185	.142	.138	1.309	.194
	Reminders	.375	.151	.275	2.489	.015**
	Security	.019	.150	.014	.127	.899

** significant at the 0.05 level

are to each other, with a scale between -1 and 1. A coefficient of 1 would indicate a perfect linear relationship, with a coefficient of 0 indicating no relationship.

The correlation check between reminders and rating average shows a statistically significant, yet moderate .351 correlation (Figure 6), roughly translating to 35% of the variance in rating average being explainable by the presence of reminders.

A new hypothesis was tested, that perhaps a score of the overall sum of positive criteria would give some indication of the mobile developers attention to overall app

Figure 6 – Correlation of rating average with reminders

		Rating Average	Reminders?
Rating Average	Pearson Correlation	1	.351**
	Sig. (2-tailed)		.000
	N	100	100
Reminders	Pearson Correlation	.351**	1
	Sig. (2-tailed)	.000	
	N	100	100

**. Correlation is significant at the 0.01 level (2-tailed).

quality, and thus have an impact on app rating average. Again, a correlation check of overall score and rating average showed a statistically significant, yet modest correlation (.252), roughly translating to 25% of the variance in rating average being explainable by the number of total criteria present in the mobile application (Figure 7).

An additional linear regression model was used, eliminating test cases with a minimum number of ratings (< 50). The theory behind this was that applications with too few ratings might skew results, as they would be more volatile to ratings and possibly more susceptible to ratings manipulation. Applications with more ratings would theoretically have rating averages that would aggregate out some of these possible impacts. This regression model fared similarly (Figure 8), with the only noticeable

Figure 7 – Correlation of rating average to total criteria present

		Rating Average	Total Att
Rating Average	Pearson Correlation	1	.252*
	Sig. (2-tailed)		.016
	N	100	91
Total Crit	Pearson Correlation	.252*	1
	Sig. (2-tailed)	.016	
	N	91	91

change being that presence of reminders no longer showed a statistically significant impact on rating average.

Yet another version of the regression model was run, eliminating applications with ratings lower than a 25% threshold (< 3.25). The argument for this test was that particularly poor ratings of applications might be due to substantial technical issues with the application, rather than the core design details of the self-tracking mobile application. Again, this model performed no better than the original, possibly a victim of the shrinking of the sample size. None of the variables met the 95% confidence level for statistical significance.

A final version of the regression model was run, testing the original model of eight criteria for only the applications concerned with weight or exercise. Through the course of coding the applications for this study, it became clear that applications designed for certain focal areas of tracking had generally different elements than applications for other focal areas. This model was used to test weight and exercise applications, as they were similar focal areas of tracking with the highest number of applications studied.

Figure 9 shows the result of this model – similarly to the original model, seven of the eight criteria did not prove to be statistically significant. In contrast to the first model, the one criteria that was a statistically significant predictor of overall rating average at a 95% level of confidence, was the ability to share, rather than the previous presence of notifications/reminders. This suggests that there are measurable differences in the impact of design criteria across different focal areas of tracking.

Figure 8 – Regression analysis (ratings > 50)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.383[a]	.147	.063	.636087324949427

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.318	.169		19.593	.000
	Early Registration	-.091	.177	-.063	-.515	.608
	Goal Setting	-.053	.146	-.041	-.365	.716
	Visualization	.209	.204	.128	1.025	.308
	Analysis	.124	.175	.081	.706	.482
	Export	.009	.142	.007	.065	.948
	Sharing	.254	.150	.191	1.691	.095
	Reminders	.270	.161	.201	1.680	.097
	Security	.030	.161	.023	.184	.855

(All statistical analysis was run through SPSS at the Odum Institute at UNC. This statistical tool was helpful not only because it allowed for regression and correlation analysis, but also because it easily allowed for filtering of records in the cases where only a subset of the sample was tested – ex. applications with ratings over 3.25.)

5. <u>Discussion</u>

While the overall predictive model of user ratings was not statistically significant, there were still things learned in this study. At the simplest level, the sample provides a picture of what types of mobile applications are currently on the market. Exercise and weight-related applications made up more than half of the sample size, and not surprisingly much of the previous research on self-tracking has been in this field. There were a large variety of other subjects tracked, including a rudimentary “meta-tracker”, that allows users to determine what data to collect on any aspect of their life.

Figure 9 – Regression analysis (weight and exercise applications)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.517[a]	.267	.139	.638798626652136

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.827	.269		10.504	.000
	Early Registration	-.061	.208	-.044	-.293	.770
	Goal Setting	.214	.197	.151	1.087	.283
	Visualization	.429	.293	.233	1.468	.149
	Analysis	.329	.314	.139	1.047	.300
	Export	-.059	.195	-.043	-.304	.763
	Sharing	.437	.204	.319	2.138	.038**
	Reminders	.342	.223	.213	1.535	.132
	Security	-.094	.215	-.068	-.435	.665

** significant at the 0.05 level

The data also provides information on what features are present in self-tracking mobile applications – with the one variable that showed significant impact on user ratings, Reminders and Notifications, showing up in only 38% of apps. The data also shows distinct differences in the measured criteria across different tracking focuses, such as the presence of goal setting in 64% of weight and exercise application, and only 43% presence in other applications.

The data also shows that average user ratings increase as expected with 7 of the 8 measured criteria. Perhaps the correlation seen between total number of criteria present and user rating can explain this – mobile application developers that take care enough to include the features can expect better user ratings.

Finally, the establishment of notifications and reminders as a significant predictor of user ratings may illustrate how users benefit from self-tracking mobile applications. Not only does the convenience of carrying mobile devices help with tracking, but the ability of the device to regularly remind users to continue their data collection seems to be a welcome feature for users.

6. Limitations of This Research/Implications for Future Research

While the analysis of the mobile application data showed some statistically significant findings for individual criteria, the overall predictive model of user ratings did not perform as well as was hoped. There are multiple possibilities for why this was the case.

Measures: The use of rating average as a gauge of user response to self-tracking mobile applications may be flawed. Ratings may be subject to manipulation, either by parties seeking to artificially improve the overall rating, or parties seeking to artificially hurt the overall rating. The specifics of when an application is rated may skew the results, as some applications provide reminders for ratings for multiple time users, while others do not. More generally, users rating an application negatively or positively may be independent of whether they choose to continue to use the application. Access to sales and use data for mobile applications would add to the understanding of how mobile applications are received. In addition, direct survey of self-tracking mobile application users attitudes may be an ideal way of measuring user response – unfortunately time and logistically made this unrealizable for this study.

Similarly the Boolean measures for the eight criteria of interest may have been flawed. Graded scales that capture more detail of the presence of specific criteria may have been more illustrative. The choice was made to use simple Boolean measures to maximize the reliability of coding decisions - creation of rating scales for the criteria, or use of existing scales might have served the study better. Laboratory user-interface testing almost assuredly would have yielded more detailed information on the role of the criteria of interest on user responses to these mobile applications – again time and logistics made such testing undoable for this study.

Criteria of Interest: Beyond possible failings in the measures used, the criteria of study used for the initial model, may have been flawed. The review of literature on mobile application development and self-tracking research independently suggested that each criteria might play a role in user response to the mobile application, but these criteria may not all have significant impact on self-tracking mobile applications. Other features external to the eight criteria in this study may have more of an impact on user response, for instance price. The choice was made to only study free applications in this study to control for the impact of varying price on users ratings – a model studying pay applications and/or using price measures for prediction may improve on analysis.

Statistical Modeling: Linear regression analysis was used to try to capture the independent effects of each of the eight criteria of interest. A multitude of other statistical methods are available, and perhaps other patterns may be found in the data collected that were missed in the regression modeling. For seven of the eight criteria of interest, the mean rating average of self-tracking mobile applications moved in the predicted direction with the presence or absence of the specific criteria. This suggests that some impact is being seen – perhaps study of a large sample of self-tracking mobile applications might show more significant findings.

7. **Conclusion**

This study is a first step in trying to understand self-tracking in mobile applications. The hope remains to delve further into the details of data analysis that users perform on self-tracking data, and ultimately, to gain an understanding of what impact self-tracking is having on people's lives.

Are self-tracking practitioners moving towards Vannevar Bush's vision of an "intimate supplement" to our memories, or are self-trackers simply adding to the "mountain of research" without analysis of their data? David Rowan speaks of the enormous potential of mining our personal data, while Sellen and Whitaker argue that the data being captured "may be less useful than we might first think" – are they both right? Have self-trackers adopted the data mining practices that companies have been using for decades, towards finding patterns in their data that "otherwise would likely remain undiscovered", or is that a future step for self-tracking? Answers to these questions and others raised in this study have implications for developers of self-tracking tools as they make their design decisions. At the same time, they have implications for the users of self-tracking tools as they seek help in their quest to better remember or better understand the patterns of their lives.

References

Allport, Gordon W. (1942). *The Use of Personal Documents in Psychological Science.* New York: Social Science Research Council.

Bederson, B.B., Teevan, J., Jones, W., (2006). Personal Information Management. *Communications of the ACM*, 49(1-5), 40.

Blair, Ann (November 28, 2010). Information overload, the early years. *The Boston Globe*

Burke, L., Wang, J., Sevick, M.A. (2011) Self-Montoring in Weight Loss: A Systematic Review of the Literature. *Journal of the American Dietetic Association*, 111, 92-102.

Bush, Vannevar (1945). As We May Think. *The Atlantic Monthly*. July 1945. 176(1), 101-108.

Cushing, C. C., Jensen, C. D., & Steele, R. G. (2011). An evaluation of a personal electronic device to enhance self-monitoring adherence in a pediatric weight management program using a multiple baseline design. *Journal of pediatric psychology*, 36(3), 301.

Data Mining. (n.d.). In *Wikipedia*. Retrieved November 23, 2011, from http://en.wikipedia.org/wiki/Data_mining.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17(3), 37-54.

Gantz, J., & Reinsel, D. (2010). The Digital Universe Decade, Are You Ready? IDC (Vol. 2009, pp. 1–16).

Germann, J. N., Kirschenbaum, D. S., & Rich, B. H. (2006). Child and parental self-monitoring as determinants of success in the treatment of morbid obesity in low-income minority children. *Journal of pediatric psychology*, 32(1), 111.

Giarelli, E., Bernhardt, B. A., & Pyeritz, R. E. (2009). Self-Surveillance by Adolescents and Young Adults Transitioning to Self-Management of a Chronic Genetic Disorder. *Health Education & Behavior*, 37, 133-150.

Hill, Kashmir (2011). *Adventures in Self-Surveillance: Fitbit, Tracking my Movement and Sleep.* Forbes, Retrieved November 30, 2001 from http://www.forbes.com/sites/kashmirhill/2011/02/25/adventures-in-self-surveillance-fitbit-tracking-my-movement-and-sleep/

IDC. (2007). *A Forecast of Worldwide Information Growth through 2010.* Framingham, MA: Gantz, John F.

Jacelon, C. S., & Imperio, K. (2005). Participant Diaries as a Source of Data in Research With Older Adults. *Qualitative Health Research*, 15(7), 991-997.

Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: The day reconstruction method. *Science*, 306(5702), 1776.

PewResearchCenter. (2011) *The social life of health information, 2011.* Washington, DC: Fox, Susannah.

Rowan, D. (2011). Personal data mining to improve your cognitive toolkit. *Wired,* retrieved from http://www.wired.co.uk/news/archive/2011-01/18/edge-question

Sellen, A. J., & Whittaker, S. (2010). Beyond total capture. *Communications of the ACM*, 53, 70.

Wilde, M. H., & Garvin, S. (2007). A concept analysis of self-monitoring. *Journal of Advanced Nursing*, 57(3), 339–350.

Yau, N., & Schneider, J. (2009). Self-Surveillance. *Bulletin of the American Society for Information Science and Technology*, *35(5), 24–30.*

Zimmerman D. & Wieder D. (1977). The diary: Diary-interview method, *Urban Life*, 5(4), 479-499.

APPENDIX I – Data Dictionary

Variable	Type	Description
AppID	autoid	Mobile App ID
AppName	Text	Registered Name of App with Apple Store
Category	Text	General Category of App
Rating Average	Numeric	Average rating of App
EarlyReg?	Boolean	Does app require upfront registration for use of prog 1=Yes, 0=No
Goals?	Boolean	Does app provide ability to set goals? 1=Yes, 0=No
Visualization?	Boolean	Does app use visualization to convey data? 1=Yes,
Analysis?	Boolean	Is any other analysis besides visualization conveyed the data? 1=Yes, 0=No
Export?	Boolean	Does app provide way to export data or access data external device? 1=Yes, 0=No
Sharing?	Boolean	Does app provide ability to share data with others? 1=Yes, 0=No
Reminders?	Boolean	Does app provide ability to set reminders for oneself 1=Yes, 0=No
Security?	Boolean	Does app offer security controls, password protection data? 1=Yes, 0=No
Total	Number	Total number of variables present in app 1=Yes, 0=

APPENDIX II – Mobile Application Reference Data

ID	AppName	Category	Rating Average	EarlyReg?	Goals?	Visualization?	Analysis?	Export?	Sharing?	Reminders?	Security?	Total
1	Calorie Counter & Diet Tracker	weight	4.563	1	1	1	0	1	1	1	1	7
2	IMapMyRun	exercise	3.868	1	0	1	0	1	1	0	1	5
3	LoseIt	weight	3.619	0	1	1	0	1	1	1	1	6
4	Period Diary	menstruation	4.307	0	0	1	1	1	0	1	1	5
5	Pedometer FREE	exercise	3.633	0	1	1	0	1	1	0	0	4
6	Calorie Counter	weight	4.159	1	1	1	1	0	1	0	0	5
7	Period Tracker Lite	menstruation	3.85	0	0	1	1	1	0	1	1	5
8	RunKeeper	exercise	3.722	0	0	1	0	0	1	0	0	2
9	Pedometer Pro Step	exercise	2.647	0	0	1	0	1	1	0	0	3
10	Pink Pad Free	menstruation	4.202	0	0	1	1	0	0	1	1	4
11	Calorie Counter: dietsÉ	weight	4.205	1	1	1	0	1	1	1	1	7
12	Stress Tracker	stress	4.769	1	0	1	1	1	0	1	1	6
13	Diet & Food Tracker	weight	2.95	0	1	1	0	1	0	0	1	4
14	iPeriod Free	menstruation	3.4	0	0	1	1	0	1	1	1	5
15	Target Weight for Adults	weight	3.011	0	1	1	0	1	0	0	1	4
16	Endomondo Sports	exercise	4.486	0	1	1	0	0	1	0	0	3
17	BabyBump Pregnancy	pregnancy	4.114	0	0	1	0	1	1	1	1	5
18	Body Fitness Free	exercise	3.03	0	1	1	0	0	0	0	0	2
19	Calorie Counter * Free	weight	3.8	1	1	1	0	0	1	0	1	5
20	Calorie Counter by MyNetDiary	weight	3.216	0	1	1	1	1	1	0	0	5
21	iDream	dream	3.756	0	0	0	1	0	0	0	1	2
22	Menstrual Calendar	menstruation	3.254	0	0	1	1	1	0	1	0	4
23	Blood Pressure Monitor	blood pressure	4.76	0	0	1	0	0	0	1	0	2
24	RxMindMe Prescription	medication	4.597	0	0	0	0	1	0	1	1	3
25	iBabyLog	parenting	4.294	0	0	1	0	0	1	1	0	3
26	Diabetes Buddy	diabetes	3.983	0	1	1	0	1	1	1	1	6
27	Blood Pressure Companion	blood pressure	4.656	0	1	1	0	1	0	1	1	5
28	Baby Nursing / Breast Feeding	pregnancy	3.163	0	0	0	0	0	0	0	0	0
29	Pregnancy Tracker Free	pregnancy	3.441	0	0	0	0	0	0	0	1	1
30	Allergy Advisor	allergies	4.417	0	0	1	1	0	0	1	1	4
31	Eat Sleep: Simple	parenting	3.594	0	0	0	0	0	0	0	0	0
32	Cancer.Net Mobile	cancer	3.84	0	0	0	0	0	0	0	1	1
33	Mint.Com	financial	3.931	1	1	1	1	1	0	1	1	7
34	Ispending	financial	4.639	0	0	1	0	0	0	0	0	1
35	Smart AlarmClock	sleep	3.668	0	0	1	1	0	0	1	0	3

APPENDIX II – Mobile Application Reference Data (cont.)

ID	AppName	Category	Rating Average	EarlyReg?	Goals?	Visualization?	Analysis?	Export?	Sharing?	Reminders?	Security?	Total
36	similac baby journal	parenting	3.413	0	0	1	0	1	0	0	0	2
37	Calorie Counter by FatSecret	weight	3.295	1	1	1	0	1	0	0	0	4
38	Pedometer Free GPS+	exercise	4.219	0	0	1	0	1	1	1	0	4
39	Fitter Fitness Calculator	weight	4.548	0	1	1	0	0	1	0	0	3
40	DailyBurn Tracker	weight	4.19	1	1	1	0	0	0	1	1	5
41	Calorie Tracker Lite	weight	4.667	0	1	1	0	0	1	1	0	4
42	FitBit Activity and Calorie	exercise	3.699	1	1	1	0	1	0	0	0	4
43	JEFIT	exercise	4.308	1	1	1	0	1	0	0	0	4
44	Bodybuilding.com	exercise	3.148	1	1	1	0	0	1	0	1	5
45	Atkins Carb Counter	weight	1.914	0	1	0	0	0	0	0	0	1
46	MyFit Fitness Workouts	exercise	3.069	1	1	0	0	0	1	0	0	3
47	Restaurant Nutrition	weight	2.776	0	1	0	0	0	0	0	0	1
48	Baby Tracker / WhattoExpect	parenting	3.379	0	0	0	0	0	1	0	0	1
49	My Days - Period & Ovulation	menstruation	4.842	0	0	1	1	1	1	0	1	5
50	Strava Cycling	exercise	4.687	0	0	1	0	0	1	0	1	3
51	Period Plus	menstruation	2.713	0	0	1	1	0	0	1	1	4
52	Pedometer Ultimate GPS	exercise	4.45	0	0	1	0	1	1	1	0	4
53	iPregnant Pregancy Tracker	pregnancy	4.1	0	0	1	0	0	1	1	0	3
54	Monthly Cycles Free	menstruation	3.891	0	0	1	1	0	0	1	1	4
55	pedometer step counter	exercise	3.457	0	0	1	0	1	1	0	1	4
56	runtastic GPS	exercise	4.015	1	0	1	0	0	1	0	1	4
57	Fitness Pro	exercise	3.19	0	0	1	0	1	0	0	0	2
58	Weight Tracker	weight	2.894	1	1	1	0	0	0	0	0	3
59	Pregnancy * Sprout Lite	pregnancy	4.659	0	0	0	0	0	0	0	0	0
60	Fertility Friend Mobile	menstruation	3.425	1	0	1	1	1	0	0	1	5
61	WomanLog Calendar	menstruation	4.491	0	0	1	1	1	0	0	1	4
62	miCoach	exercise	3.66	1	1	1	1	1	0	0	1	6
63	FitDay Mobile	weight	2.667	1	0	1	0	1	0	0	1	4
64	GoMeals	weight	2.761	0	1	0	0	0	0	0	0	1
65	Sports Tracker	exercise	4.572	0	0	1	0	1	1	0	0	3
66	Joggy Coach Free	exercise	2.379	0	0	1	0	0	1	0	0	2
67	Free Menstrual Calendar	menstruation	3.12	0	0	0	1	1	0	0	1	3
68	Women's Health Lite	exercise	3.488	0	1	0	0	1	0	0	0	2
69	I'mExpecting	pregnancy	3.597	0	0	0	0	1	1	0	0	2
70	Waterlogged	water intake	4.209	0	1	1	0	0	1	1	0	4

APPENDIX II – Mobile Application Reference Data (cont.)

ID	AppName	Category	Rating Average	EarlyReg?	Goals?	Visualization?	Analysis?	Export?	Sharing?	Reminders?	Security?	Total
71	Glucose Buddy	diabetes	3.725	0	0	1	1	1	0	1	0	4
72	iRunner	exercise	3.82	0	0	1	0	1	1	0	0	3
73	Period Log Free	menstruation	4.553	0	0	1	1	1	0	1	1	5
74	Strava Run	exercise	4.781	1	1	1	0	0	1	0	1	5
75	StrongLifts	exercise	4.73	0	1	0	1	0	0	0	0	2
76	WiScale	weight	2.771	1	0	1	0	0	1	1	1	5
77	Workout Logger	exercise	2.767	0	1	1	0	1	0	0	0	3
78	Pedometer Multifunctional Free	exercise	3.6	0	0	1	0	1	0	0	0	2
79	iMapMyRIDE - Cycling	exercise	3.352	1	0	1	1	1	1	0	1	6
80	Sleep Tracker	sleep	2.855	0	0	1	0	0	0	0	0	1
81	Pregnancy Lite	pregnancy	3.727	1	0	1	0	0	1	0	0	3
82	t2 Mood Tracker	mood	4.119	0	0	1	0	0	0	1	1	3
83	Track & Share	all types	2.777	0	1	1	0	1	1	1	1	6
84	MoodPanda	mood	2.722	1	0	1	0	0	1	0	1	4
85	WaveSense Diabetes Manager	diabetes	3.377	0	1	1	0	1	0	0	0	3
86	Diabetes Companion	diabetes	3.181	0	1	1	0	1	0	0	0	3
87	Diabetes Log	diabetes	2.9	0	0	0	0	1	0	0	0	1
88	Running Log Free	exercise	3.221	0	0	0	0	0	0	0	0	0
89	LogYourRun Lite	exercise	3.783	1	0	1	0	0	1	0	0	3
90	atimeLogger	time management	4.492	0	1	1	0	1	0	1	0	4
91	Calorific	weight	3.601	0	1	1	0	0	0	1	0	3
92	MapMyFitness	exercise	3.899	1	1	1	0	0	1	0	1	5
93	FitnessBuilder	exercise	3.588	1	1	1	0	1	1	1	1	7
94	Period and Fertility Tracker	menstruation	4.098	0	0	1	1	1	1	1	1	6
95	Map My Tracks OutFront	exercise	3.315	1	0	1	0	1	1	1	1	6
96	Footsteps Pedometer Free	exercise	2.895	0	0	0	0	0	0	0	0	0
97	myPlan - Special K Challenge	weight	3.243	1	1	1	0	0	0	1	1	5
98	Target Weight for Teens	weight	3.186	0	1	1	0	1	0	0	1	4
99	Nutrition Genius Free	weight	2.733	0	1	0	0	0	0	0	0	1
100	Monitor Your Weight	weight	4.46	0	1	1	0	1	0	1	1	5

ACKNOWLEDGEMENTS

I would like to thank Dr. Arcot Rajasekar for his crucial advice and patience throughout this project. I would also like to thank the Odum Institute for assistance with statistical analysis. Finally, I am grateful to my family and friends who have endured me during the last year of this process.