Information Artifact Evaluation and Iterative Design:
A Novel Mixed-Method Approach

Timothy S. Carlson

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2017

Reading Committee:
Hans J. Scholl (Chair)
Michael B. Eisenberg
Megan Finn
Jennifer Lee Hoffman
Hala P Annabi (Reader)

Program Authorized to Offer Degree
Information School

University of Washington

Abstract

Information Artifact Evaluation and Iterative Design:
A Novel Mixed-Method Approach

Timothy S. Carlson

Chair of the Supervisory Committee:
Hans J. Scholl, Professor
The Information School

Organizations face numerous design challenges regarding effectiveness and measurability of success when investing in Information Systems and Information Technology (IS/IT), or information artifacts for short. Some information artifacts produce better informational and transactional results than others, but it is not clearly understood which information artifacts do better, and why. The designing, building, maintaining, and upgrading of information artifacts needs to be better informed, more targeted, and predictably effective. This requires a deeper

understanding of context-sensitive artifact evaluation criteria and their link to design choices. Combining qualitative and quantitative evaluation and outcome testing methods promises to be an innovative way to improve artifacts iteratively.

The contribution of this study is twofold: It provides an empirical examination and validation of an iterative and phased assessment approach, and it introduces a novel methodological approach. Through empirical testing, qualification, and validation, the results demonstrate that information artifacts can be targeted and improve. This has great relevance to information artifact evaluation and design practice in any organization of any size and any sector.

The study employs the TEDS assessment framework and assessment methodology to evaluate and rank 10 iOS mobile sports applications in eight use case scenarios. A design method is developed to leverage the TEDS rater output of screenshots and comments from the evaluations. Through brainstorming, design patterns are identified based on the highest performing assessed applications and used to assist directive treatment development. Feature treatments are created and used in A/B testing by the organization of one of the 10 assessed iOS mobile applications. Upon completion of the A/B testing phase and treatment deployment, a subsequent TEDS evaluation of the application was performed that showed improved ratings for the application.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to my advisor and committee chair, Dr. Hans Jochen Scholl. He has provided insight, challenge, and encouragement through this long PhD journey. Dr. Scholl is a brilliant researcher, yet his generosity and kindness have impacted me just as much. Without his contribution, this dissertation would not have been possible. He is the ideal mentor and someone I will always aspire to emulate.

I would like to thank my committee members, Dr. Michael B. Eisenberg, Dr. Megan Finn, Dr. Jennifer Lee Hoffman, and Dr. Hala P. Annabi. Your willingness to dig into this study and provide such valuable insight has enriched this study and assisted my own learning.

I would like to recognize Lee Dirks, whom Dr. Scholl, Dr. Eisenberg and I had the pleasure of working with on the TEDS framework. What the three of us would give to have a chance to work with Lee again.

I would like to thank Dr. Hazel Taylor, Dr. Michelle Carter, and Dr. James Antony for their contributions while members of my committee. I am very fortunate to have benefited from your contributions while our paths crossed at the University of Washington.

I would like to thank Tosh Meston, James Milgate, and Chris Bryant from the Sounders FC. I would like to thank the other TEDS researchers who have contributed greatly to this study, Grant Woods, Dustin Chiang, Mario Sanchez, and William Menten-Weil.

I would like to thank Don Crowe, Jeff Grove, Phil Fawcett, Terry Marks, and Teresa Taylor, who all have contributed in various ways to this academic journey.

DEDICATION

I dedicate this dissertation to my wife Betsy, who has patiently supported me on this long trek. I love you Betsy. Thank you to Peter, Erik, and Kevin for understanding those times when I have been buried under a stack of journal articles. Spending time with you all is precious to me. I am so thankful for you, my family.

Chapter 1

INTRODUCTION

Background

Small and medium enterprises (SMEs) face challenges determining how to best invest in Information Systems and Information Technology (IS/IT). Adoption of IT tends to be slow for these businesses, and once adopted, the failure rate is high (Nguyen et al., 2015). The adoption gap of IS/IT between SMEs and larger firms is significant, with evidence suggesting that SMEs are not taking full advantage of the digital opportunities (Morgan-Thomas, 2016). Yet, as of 2013, 74.4 percent of households in the US had a desktop, laptop, or handheld computer with some kind of Internet subscription; 73.4 percent of households had high-speed Internet access (File & Ryan, 2014). Over 90 percent of US households have a mobile phone and over 33 percent have a smartphone (Kaplan, 2012). Internet access has become ubiquitous.

The assessment and evaluation of IT investments is an important aspect of determining whether they are helpful. This dissertation lives in the intersection between industry and the academy. The past decade has brought advances in evaluation methods of IT artifacts in both academic literature and in industry practice. An opportunity exists to leverage advancements in both domains to make the building, maintaining, and upgrading of IT investments on the web more effective and efficient.

In academic literature, a variety of methods have emerged to evaluate IT artifacts. These methods come from two perspectives. The first perspective takes a system-centric approach that views assessment through the lens of user acceptance of a technology. This perspective asks: What are the things that motivate the usage and acceptance of a system? These methods are referred to as *"user acceptance frameworks."* The second approach takes a user-centered perspective, using a set of user criteria (categories) to ask: How much satisfaction does a user get from each criterion? These methods are referred to as *"evaluation and assessment frameworks."* Although both methods provide insight about the strengths and weaknesses of an information system, neither method provides a step-by-step approach for improving a system. This study makes a contribution by providing a step-by-step method.

In the software industry, leading companies like Google, Amazon, and Microsoft use A/B testing, which is a randomized experiment that uses two variants (A and B), such as two versions of a webpage, to evaluate whether proposed changes achieve the desired impact. Done correctly, A/B testing provides scientifically valid experiments by using data-mined system metrics to make this determination. Although A/B testing does not help an organization determine the features to test nor how to evaluate the priority of features in their feature backlog, it gives them a methodology to evaluate the merit of different proposed versions of a feature based on customer behavior. A/B testing is gaining popularity in industry. A company can create a competitive advantage by using an assessment framework to identify weaknesses in a website, to create treatments, and then to conduct A/B testing to determine whether the treatments improve user-driven system metrics.

In this document, *information artifact* refers to sources of information, pieces of information, information systems, and other information technology artifacts (Scholl & Carlson, 2012). This summary term recognizes that information in its many constructs and diverse technology instantiations can no longer be meaningfully distinguished. For example, a human-readable XML document with embedded code and self-executable functionality can be considered a meaningful piece of information or an information technology artifact and information system—or both at the same time (Scholl et al., 2011a).

Problem Statement

An online presence is fundamental to business today, yet many organizations lack the expertise to evaluate and improve their website and related information artifacts. While SMEs lag larger organizations in adopting IS/IT advancements (Morgan-Thomas, 2016), even large organizations with in-house expertise skip evaluations because they are time-consuming and costly (Scholl & Carlson, 2012). A/B testing can be a valuable quantitative method to provide organizations confidence that proposed changes make a positive impact for their websites and artifacts, yet there is little guidance to help organizations identify what to test. In addition, A/B testing provide quantitative metrics without explaining why something is better. Understanding how to assess whether a company website is useful and usable is complicated, with many factors to consider. Plus, usefulness and usability can vary for a website, depending on the user task. Even when employing an assessment, it is a challenge to generate targeted feature changes that address weaknesses found and quickly implement updates to address them. Organizations need a research-validated method to evaluate and improve their information artifacts in a timely and cost-effective manner. On its own, A/B testing is not sufficient because it can only identify which version of a feature performs better. It cannot identify which features need to be improved, nor determine changes that might improve the artifact. Assessment and evaluation frameworks consider these factors, but they do not provide a way to quantitatively validate whether a proposed change has a positive impact based on user generated metrics.

Choosing an Assessment Framework

The literature review discusses numerous noteworthy evaluation and assessment frameworks. The TEDS assessment framework (Scholl et al., 2011b) was chosen for this study for two reasons: it includes a comprehensive list of sub-criteria, and its proven 13-step assessment procedure for gathering rater-generated comments and screenshots, which can be used to help generate feature improvement ideas for assessed artifacts.

Purpose Statement

The purpose of this research is to study the effectiveness and efficiency of an assessment approach to developing web-based information artifacts based on the TEDS assessment framework, iterative feature analysis and modification, and A/B testing. This study will provide insights into the nature and usefulness of the TEDS model for assessments that seek to generate targeted feature modification treatments using insights from the TEDS scenario-based evaluator comments and screenshots. This study will also demonstrate that A/B testing provides quantitative experimentation data that informs the development team whether a proposed feature measurably improves the information artifact. It will demonstrate the integration of TEDS evaluations, targeted treatment improvement design specifications, and A/B testing in a timely and cost-efficient manner with a leading sports iOS mobile sports application.

Significance of the Study

The study provides the hands-on use of a robust academic assessment framework along with A/B testing in an iterative fashion. For small and medium-sized enterprises (SMEs) and even large businesses, this study can serve as a template for time- and cost-efficient evaluations, assessments, comparisons, and targeted design recommendations. For IT companies that regularly use A/B testing, this study shows them how they can gain finer-grained insight into which features they should A/B test by iteratively applying the TEDS assessment framework in combination with their A/B testing.

The artifacts evaluated in this study are mobile sports applications for the iOS (Apple) platform, thus demonstrating the applicability of this method for organizations who build web-based sports-related information artifacts. However, the approach is not limited to sports-related information artifacts. It can be applied to any information technology domain by organizations of any size.

This dissertation addresses the identified problem in three phases:

Phase 1: Assessment - Previous studies have shown that using an assessment framework to evaluate and compare websites provides insight into their relative strengths and weaknesses (Scholl & Carlson, 2012; H. J. Scholl et al., 2014). Although using an assessment framework can provide interesting explanatory information about the performance of an information artifact service, the information is only useful if it leads to changes that improve the service. The assessment ratings and other evaluator-generated feedback (like comments and screenshots) are structured to enable the design idea generation phase to be focused and specific. It is even more helpful if the framework provides examples of other information artifact services that perform well along the same assessment criteria.

Phase 2: Design idea generation - To increase an organization's motivation to improve the information artifact rather than simply evaluate it, a phase to generate design recommendations is necessary. During Phase 2, the organization generates feature additions or recommendations to improve the information artifact service.

Phase 3: A/B testing - With design recommendations in hand, an organization needs a safe method to evaluate whether the recommendations are effective. This third phase provides insight into the value of the proposed design recommendations. A/B experimentation offers a scientifically rigorous method for evaluating feature changes. Using their existing information artifact service, organizations can test design ideas using a subset of their users in a controlled manner. Using data mining techniques, they can determine if the behavior of those users shows the desired benefits in comparison to the behavior of users using the existing information artifact service.

Subsequent studies that use the same framework to compare the same websites at a later date have shown that organizations must continually improve and add features to compete and to meet the needs of their existing and potential users (H. J. Scholl et al., 2014). By using an iterative method during assessment, design recommendations, and design recommendation evaluations, organizations can address comparative weaknesses in their information artifact services now and as they evolve.

Assumptions and Limitations

The TEDS assessment framework has been implemented multiple times during the past six years to evaluate websites and applications. (See Appendix A for TEDS Framework details.) The results are encouraging: In each instance, the TEDS framework provided results that made it easy to identify the strengths and weaknesses of websites and apps. However, longitudinal analysis of the previous TEDS assessments has shown that the features and competitive landscape of websites and apps are dynamic and thus the TEDS assessments are time-sensitive (Scholl & Carlson, 2012). Given the evolving nature of technology on the Internet, if these organizations lack the resources or motivation to address issues in a timely manner, the TEDS assessment may need to be re-run to provide the most relevant evaluations.

The implication for this study was that the TEDS assessment phase and A/B testing phase needed to be performed within a reasonable timeframe based on the dynamic nature of the environment. TEDS had been used to evaluate sports websites, which provide sports news, game scores, team standings, and team schedules, among other related sports information (Jurisch et al., 2014) (Scholl, 2013) (Scholl & Carlson, 2012). These studies showed significant evolution of evaluated websites within months of initial assessments. To minimize this risk, this study was planned to occur within one soccer season for the Seattle Sounders FC (football club), which is roughly nine months.

The TEDS framework can be used to assess almost any information artifact, and it can be used with a diverse set of scenarios. This research is limited to a web-based study using a data analytics toolset that supports A/B testing. Several data analytic platforms are capable of A/B testing, such as Google Analytics or Yahoo Flurry Mobile Developer Suite. Regardless of the data analytics platform that is chosen, A/B testing requires adding java script to webpages to support the analytics tracking.

Most mobile scenarios can be A/B tested, since mobile features are typically implemented as normal webpages. However, features that reside on an external partner platform, yet are integrated into the application through hyperlinks or iframes, were not considered for inclusion in this study. For example, the webpage of a soccer club that uses Facebook social media features does not have access to Facebook's platform to create treatment pages on Facebook for A/B testing. Their exclusion was due to the difficulty of creating A/B treatment pages on another platform and subsequently getting access to the data mined from Facebook for the A/B analysis.

This study relied on the willingness of one or more participating professional soccer clubs to use the results from the initial TEDS assessment to A/B test proposed features. To run the A/B experiments, the club needed access to the source code of the features and pages being tested. Treatment pages and features needed to be created, along with small amounts of code written. Many clubs outsource parts of their websites, which limits access to pages and features.

The Seattle Sounders FC agreed to participate in this study through the treatment development phase and A/B testing phase. This was an optimal partnership because the development team for the Seattle Sounders FC is near the University of Washington, making it convenient for face-to-face meetings and collaboration.

This study was not done in a controlled laboratory environment. In a laboratory environment, the researcher has control over many of the processes and can limit outside variables. In this study, adjustments needed to be made to address requirements and changes in a live production development environment. For example, the Sounders FC development team had competing development priorities and development resource limitations that delayed their development work for this project and limited their choice of A/B testing platform because of the development work required to implement it. Thus, adjustments were made to the project to reduce the scope of the treatments that were developed and to choose an A/B testing platform that required less development effort by the Sounders FC development team. These adjustments allowed the project to move forward and meet the research goals with only a slight delay. This flexibility lends credibility to this study because it demonstrates that it is relevant in development organizations. (See Appendix A for the terms and definitions used in the TEDS assessment framework.)

Chapter 2

LITERATURE REVIEW

This literature review contains seven sections: the theoretical foundations of assessment frameworks found in academic literature; common characteristics found in those frameworks; the TEDS value criteria, a comparison of the frameworks with the TEDS framework; data mining and web analytic methods for analyzing and comparing information artifacts; A/B testing as a methodology for evaluating feature changes to improve a website; and finally, the application of online A/B testing with the TEDS framework.

Theoretical Foundations of Assessment Frameworks

High-tech information systems like websites, search engines, digital car dashboards, and smart phones and smart phone apps are pervasive today. Academic literature from the past 30 years provides numerous models developed to measure the quality of information systems, resulting in different frameworks of dimensions. Although differences exist between the academic frameworks, each framework identifies sub-criteria or attributes that should be considered when evaluating an information system or artifact. Two prevalent perspectives of information artifact assessment frameworks emerge from this literature.

The first perspective that appeared in the IS literature takes a system-centric view by asking, What motivates the usage and acceptance of a system? The methods that developed from this perspective are referred to as *user acceptance frameworks*. The second perspective, which emerged in the mid-1990s, takes a user-centered perspective. It typically evaluates information artifacts using a set of user criteria that address the question: How much satisfaction does a user get from each criterion? The methods that developed from this perspective became known as *evaluation and assessment frameworks*. Table 1 shows information artifact assessment frameworks found in the literature and their primary authors. This review includes the most prolific assessment frameworks found in the literature. It also discovered frameworks that show promise for generating feature-improvement ideas that can be used with A/B testing. In the graphs in this document, theory from IS literature appears in blue. Theory from HCI literature or related field is in green. Theory from Information Science literature appears in purple.

Table 1. Information Artifact Assessment Frameworks

Perspective	Framework	Author	Literature Stream
Acceptance of Info Artifacts	Technology Acceptance Model (TAM)	Davis Chan and Teo Lederer, Maupin, Sena, Zhuang Chuan-Chuan Lin and Lu	Information Systems
Acceptance of Info Artifacts	D&M Information Success model	DeLone and McLean Seddon (modified D&M)	Information Systems
Acceptance of Info Artifacts	Task-Technology Fit (TTF)	Goodhue and Thomson	Information Systems
Acceptance of Info Artifacts	User Acceptance of Information Technology (UTAUT)	Venkatesh, Morris, Davis and Davis	Information Systems
Evaluation and Assessment	IS/SQ Disconfirmation	McKinney, Yoon, and Zahedi	Information Systems
Evaluation and Assessment	Aim Quality (AIMQ)	Lee, Strong, Kahn, and Wang	Information Systems
Evaluation and Assessment	Five Dimension System	Yang, Cai, Zhou, and Zhou	Information Systems
Evaluation and Assessment	Data Quality Dimensions (DQD) Ontology	Wand & Wang	HCI/Other
Evaluation and Assessment	WebQual	Barnes & Vidgen	HCI/Other
Evaluation and Assessment	Adaptive Decision Making (ADM)	Hartman, et al	HCI/Other
Evaluation and Assessment	Value Added	Taylor	Information Science
Evaluation and Assessment	TEDS	Scholl, Eisenberg, Dirks, Carlson	Information Science

Academic Domain	
	Information Systems
	HCI/Other
	Information Science

User Acceptance Frameworks

User acceptance frameworks seek to understand whether an organization would successfully adopt a given information artifact. These frameworks first appeared in the IS literature in the late 1980s from Fred Davis (Davis, 1989), who introduced the Technology Acceptance Model (TAM). DeLone and McLean published their D&M Information IS Success Model in the early 1990s (William H. DeLone & McLean, 1992). The fundamental concepts in these two seminal frameworks have been refined and extended over the years to remain insightful about the acceptance of information artifacts frameworks in an organization. The TAM framework bifurcates acceptance of a system into two concepts: Perceived Usefulness (PU) and Perceived Usability or Perceived Ease of Use (PEOU).

More recent assessment and evaluation frameworks appear in the IS literature; these include multiple characteristics instead of a binary acceptance criteria (Lee et al., 2002), (McKinney et al., 2002), (Yang et al., 2005). For example, the D&M model uses five categories to determine the acceptance of a system: systems quality, information quality, service quality, use, user satisfaction, and net benefit (system benefits).

User acceptance frameworks were derived from a system perspective rather than an end-user perspective. The benefit of a system perspective is that it includes constraints in a framework that are significant to an organization. For example, some features require high maintenance costs, which might outweigh the end-user benefit to the organization of building that feature.

Evaluation and Assessment Frameworks

Evaluation and assessment frameworks began to emerge in 1996 in Human Computer Interaction (HCI) and related disciplines. Whereas the earlier information artifact frameworks appeared before the web revolution, they were intended for use with conventional (non-web) systems. Because HCI-based frameworks were developed during the Internet era, they incorporated a web perspective from the outset, making them well-suited to the evaluation of web artifacts.

Unlike the user acceptance frameworks found in early IS literature, which included binary decisions of user acceptance, the evaluation and assessment frameworks consider numerous attributes, which provides a more articulated assessment. For example, an evaluation and assessment framework might rate a web-based search engine high on the look-and-feel of the user interface, but rate the results of a search low because the returned results were not relevant based on the search terms that were provided by the user.

HCI research asserts that designers of information artifacts should:

- Focus early on users and tasks
- Empirically measure designs with users using prototypes and simulations
- Design iteratively (Gould & Lewis, 1985)

HCI emphasizes that systems designed for people need to prioritize ease of use, learnability, and usefulness. HCI-based assessment mechanisms prioritize such models as Adaptive Decision Making (ADM) (Hartmann, De Angeli, et al., 2008) and the Data Quality Dimensions in

Ontological Foundations method (Wand & Wang, 1996). Table 2 presents a timeline of the assessment frameworks since 1980 that are discussed in this dissertation.

Table 2. Assessment Framework Timeline

1980	1985	1990	1995	2000	2005	2010	
Information Science							
Value Added (Taylor 1980)							TEDS (2011)
Information Systems (IS)							
		TAM (Davis 1989)			TAM (Lederer 2000)TAM (Lin 2000)		
						TAM (Chan, Teo 2007)	
			D&M IS Success (1992)				
				TTF (Goodhue and Thomson 1995)			
				D&M IS Success (Seddon 1997)			
					IS-SQ Disconfirm (2002)		
					AimQ (Lee, Strong, et al, 2002)		
					UTAUT (Venkatesh, et al, 2003)		
						5 Dimension SQ (Yang, et al, 2005)	
HCI/Other							
				DQD Ontology (Wand & Wang '96)			
					WebQual (Barnes & Vidgen 2000)		
						ADM (Hartman, et al, 2008)	

All these frameworks are worthy assessment tools that provide a unique perspective to assess information artifacts. Although they emphasize unique points and include different evaluation attributes or sub-criteria, many themes and criteria overlap, particularly in the terms that are used to describe the criteria. The frameworks also uniquely interpret the terms information quality, usefulness, usability, ease of use, and value.

The TEDS framework emerged as the best choice for this dissertation because of its comprehensive framework and the usefulness of its 13-step assessment procedure for generating feature improvement ideas for assessed artifacts. As a result, TEDS is introduced in detail in the next section. Then, the sections that follow discuss common terms, some of the nuanced differences in the perspectives of the evaluation and assessment frameworks, and a comparison of their sub-criteria and attributes.

The TEDS Framework

The TEDS framework by Scholl, Eisenberg, Dirks, and Carlson was also born out of HCI, but it takes an information-centric perspective. It builds on *value-added frameworks*, which arose in 1980 with the original Taylor model (Taylor, 1982). These value-added frameworks (including the TEDS refresh in 2011) are evaluation and assessment frameworks. They include many attributes, rather than focusing on whether an information artifact framework was accepted by a user or organization, as modeled by the user acceptance frameworks. Taylor's original framework was created before the Internet; the TEDS update added the web perspective to the value-added perspective.

The TEDS framework provides a method for describing, evaluating, and comparing information artifacts. (Page 1 in Appendix B lists Taylor's original 23 values.) TEDS extended and reorganized the Taylor's values somewhat, modified some definitions, and added new values that emerged in the past 25 years. The TEDS framework includes 40 values (or sub-criteria) that are organized into six major user criteria:

- Ease of use
- Noise reduction
- Quality
- Adaptability
- Performance
- Affection

TEDS uses the terms sub-criteria and value interchangeably. (See Appendix C for the definitions for all TEDS sub-criteria in each user criteria.) This taxonomy covers the major themes of information quality, usefulness, and usability in assessment frameworks at a very fine level of detail. The 40 sub-criteria come with some challenges, because although each item has a distinct definition and meaning, the TEDS literature has suggested that not all sub-criteria are equally important in each assessment context. Research is needed to prioritize the most relevant sub-criteria for each assessment context. Nonetheless, provided there is a way to effectively prioritize or cluster sub-criteria for a given context, the level of detail in the TEDS framework should be considered a benefit because it is possible to drill to the level of detail when desired.

TEDS also extended Taylor's original model by leveraging the concepts from HCI of using scenarios and personas to integrate the context-specific perspective into the evaluation process. Finally, the TEDS framework contributed a pragmatic 13-step procedure for evaluating information artifacts. The procedure was included in the TEDS framework to be used in empirical comparative and evaluative research (Scholl et al., 2011b).

Common Characteristics of the Frameworks

In the assessment literature of HCI, Information Systems, and Information Science, the terms information quality, usefulness, usability, and ease of use are used in different ways. This reflects the unique lens or perspective of each model. The frameworks also vary in complexity; some propose more parsimonious models, while others provide more detail. With its 40 attributes, the TEDS model offers the most extensive framework for evaluating information artifacts. However, providing more attributes does not make a framework better for an evaluation context if the framework is missing important, relevant sub-criteria, or the sub-criteria do not add unique, relevant detail. The following sections discuss commonly used concepts in the literature, and highlight the use of concepts in the TEDS framework.

Information quality

Information quality is a pervasive concept in assessment and evaluation literature, yet there is no agreement on its definition. Klischewski and Scholl identify Information Systems and Information Science as the two major literatures that contribute to information quality (IQ). They suggest that Information Systems literature emphasizes a technology perspective, and Information Science literature emphasizes an information perspective (Klischewski & Scholl, 2008). Although overlap exists in the way HCI and IS literature discuss information quality, HCI literature also contributed to the taxonomy, with the first user-focused approach to IQ. Given that IQ is a nebulous (Ballou et al., 2003), context-sensitive, and elusive concept (Lillrank, 2003), it is useful to include these various perspectives to make sense of IQ.

Many early IS frameworks use the terms information quality and system quality. The D&M Information IS Success model uses system quality as a measure of technical success, and information quality as a measure of semantic success (W. H. DeLone & McLean, 2003). Lee, Strong, Kahn and Wang divide IQ into four categories: Sound IQ, Dependable IQ, Useful IQ, and Usable IQ (Lee et al., 2002). McKinney splits IQ into Web IQ factors and Web SQ factors (McKinney et al., 2002). More recent IS evaluation and assessment frameworks also use IQ as a major category (Yang et al., 2005).

Many evaluation frameworks use other terms to describe IQ concepts. Wand and Wang include accuracy and precision, reliability, timeliness, completeness, and consistency (Wand & Wang, 1996). Goodhue and Thompson include quality, *locatability* (the ease of determining what data is available and where), compatibility, ease of use/training, timeliness, system reliability, and relationship with users in Task-Technology Fit (TTF) dimensions (Goodhue & Thompson, 1995). Hartman, De Angeli, and Sutcliffe include content, aesthetics, reputation, and customizability (Hartmann, Sutcliffe, et al., 2008). The TAM framework adds attributes, such as clear and understandable, flexible, and controllable (Davis, 1989). (See Appendix B for a side-by-side comparison of the frameworks.)

Rather than using the terms information quality or system quality, the original Taylor model included terms that represent IQ themes. The 2011 TEDS update added a user criteria quality that consists of six values: accuracy, comprehensiveness, currency, reliability, validity, and authority. (See Appendix C, Table 6 for definitions of these values.) Accuracy, currency (McKinney et al., 2002), and reliability are also seen in the IQ/IS delineation in IS frameworks (W. H. DeLone & McLean, 2003; McKinney et al., 2002).

TEDS adds comprehensiveness, validity, and authority, which are all significant values to web users. Validity and authority are especially valuable for evaluating the trustworthiness of information. The variety in the evaluation and assessment frameworks shows that IQ is context sensitive and multifaceted. The TEDS framework includes the most robust set of attributes that describe information quality. Table 3 shows the categorizations for information quality that are used in the assessment literature.

Table 3. Information Quality

Term	Reference	Description		Date
Information Quality System Quality	DeLone and McLean	*Information Quality:* • Completeness • Ease of understanding • Personalization • Relevance • Security	*Systems Quality:* • Adaptability • Availability • Reliability • Response time • Usability	1992
Sound Information Dependable Information Useful Information Usable Information	Lee, Strong, Kahn	*Sound IQ:* • Free of error • Concise representation • Completeness • Consistent representation+ *Dependable IQ:* • Timeliness • Security	*Useful IQ:* • Appropriate amount • Relevancy • Understandability • Interpretability • Objectivity *Usable IQ:* • Believability • Accessibility • Ease of operation • Reputation	2002
Web Information Quality System Information Quality	McKinney, Yoon, & Zahe	*Web IQ:* • Relevance • Timeliness • Reliability • Scope • Perceived Usefulness	*System IQ:* • Access • Usability • Navigation • Interactivity	2002
Information Quality	Yang, Cai, et al.	• Usefulness of content • Usability		2005
Information Quality	Wand and Wang	• Accuracy and Precision • Reliability • Timeliness & Currency • Completeness • Consistency		1996
Information Quality	Hartmann, De Angeli, Sutcliffe	• Usability • Content • Aesthetics • Reputation/identity • Customizability		2008
Information Quality	Scholl, Eisenberg, Dirks, & Carlson	• Accuracy • Comprehensiveness • Currency • Reliability • Validity • Authority		2011

Usefulness, usability, and ease of use

Like information quality, the terms usefulness, usability, and ease of use occur in most of the frameworks. Sometimes they are used synonymously (particularly usability and ease of use), but other times they are compared as unique variables that influence user acceptance of technology. Table 4 defines usefulness in the assessment literature.

Table 4. Usefulness

Term	Reference	Description	Date
Usefulness	Davis	• Capable of being used advantageously	1989
Perceived Usefulness	McKinney, Yoon, & Zahe	• Users' assessment of the likelihood that the information will enhance their purchasing decision • Defined in the "Web IQ" category	2002
Usefulness	Yang, Cai, et al.	• Unique content • Relevant information to the customer • Valuable tips on products/services • Reliable professional opinions • Up-to-date information	2005
Usefulness	Chan and Teo	• A system that increases a user's job performance within an organizational context	2007

Each description in the table reinforces a theme of utility and solidifies a common understanding among academics about the definition of usefulness. This understanding aligns with the Merriam Webster definition of *usefulness*: "The quality of having utility and especially practical worth or applicability."

Building on Taylor's theory that usefulness determines value as perceived by the user (Taylor, 1982), the values of an information artifact can measure whether the system is useful to the user. In this way, usefulness is synonymous with value, which is used as a currency, measurement, or even a magnitude of the user criteria expressed in the framework.

TEDS does not explicitly use the term usefulness as a criterion or value because multiple attributes contribute to something being useful. Thus, TEDS uses multiple values to represent various characteristics of usefulness. For example, TEDS includes cost savings, time savings, security, and safety from the performance criteria. (See Appendix C, Table 8 for these definitions.) It also includes linkage/referral and order from the noise reduction criteria. *Linkage/referral* refers to a user's ability to gain deeper understanding through pointers and links to resources outside of the specific information artifact. *Order* refers to the extent to which information is uncluttered and non-distracting, thereby providing clearer insight. (The definitions for noise reduction criteria are found in Appendix C, Table 9.)

Adaptability in the TEDS framework criteria includes 10 values: contextuality, flexibility, simplicity, transaction, trust, feedback, community, individualization, localization, and privacy. Affection in the framework includes five values: aesthetics, entertainment, engagement, stimulation, and satisfaction/rewarding/incenting. (See Appendix C, Table 10 and Table 11 respectively, for definitions of these values.)

The terms usability and usefulness are related but distinct: Usability refers to attributes that contribute to ease of use, whereas usefulness refers to attributes that contribute to utility: Can it get the job done? The understanding of usability differs in the literature. Many authors use usability to mean ease of use. Other authors include useful or utility in their definitions of usability. This amalgamation of concepts is unfortunate; a distinction should exist between an artifact being easy to use versus having utility. Table 5 defines usability in the assessment literature.

Table 5: Usability

Term	Reference	Description	Date
Usability	Gould and Lewis	• Functions that users really need in their work	1985
Usability	Rubin	• Usefulness (relative to task/intent) • Effectiveness (ease of use) • Learnability (ease of learning and retention) • Attitude (likeability)	1994
Usability	Ivory and Hearst	• The extent to which a computer system enables users, in a given context of use, to achieve specified goals effectively and efficiently while promoting feelings of satisfaction (ISO 1999)	2001
Usability	McKinney, Yoon, and Zahe	• The extent to which a website is visually appealing, consistent, fun, and easy to use • Usability is defined in the "System IQ" category	2002
Usability	Yang, Cai, et al.	• Well-organized hyperlinks • Adequacy of security features • Search facilities • Customized search functions • Customized information presentation • Confidentiality for customer information	2005
Usability	Hartman, Angeli, et al.	• Ease of learning • Ease of use • Memorability • Low error frequency • Subject satisfaction	2008

Like usefulness, usability consists of more than one specific user criterion or value. TEDS uses the term ease of use, and gives it seven values: browsing, formatting, mediation, orientation, order, accessibility, and simplicity. (See Appendix C, Table 7 for definitions of these values.) This enumeration makes it possible to describe subtle differences that make an information artifact more usable or less usable. It also provides the opportunity to describe something as being easy to use in some ways, while being difficult or frustrating in other ways.

Value

To evaluate information artifacts, TEDS uses the term value, IS uses the term IQ, and HCI uses the terms usefulness, usability, and ease of use. TEDS use the concepts of usefulness, usability, ease of use, aesthetics, and quality in various user criteria and values. Although they overlap with terms used in other frameworks, the value-added perspective provides a different scale of measurement.

Taylor stated that usefulness is the determination of value as perceived by the user (Taylor, 1982). This user-centered perspective was novel in the early 1980s. By the early 1990s, it had become prevalent in the HCI literature (Rubin, 1994). Value is affected when problems occur that require information resolution. Thus, information artifacts can affect value. In the TEDS framework, information systems are a series of formal processes that increase the potential usefulness of input messages processed. This builds on Taylor's assertion that "the primary reason for the existence of formal information systems is to add value to information" (Taylor, 1982) (p. 341).

Therefore, *information systems* are any formal set of value-adding processes that help data lead to knowledge, productive knowledge, or action. Taylor knew that the value of information is context-specific. An information message is given value by the individual who uses the message. Taylor suggested that information systems have historically been driven by either content or technology. Content-driven design comes from traditional divisions of knowledge that have been applied in the fields of science and other scholarly fields for more than several thousand years (Taylor, 1986)

(p. 3). Classification-based search tools, in which the user navigates a scientific taxonomy of mammals or plants to find information, are examples of content-driven design. Technology-driven design has arisen from technology as it has evolved over the years. For example, CD-ROM-based encyclopedias, which were built for personal computers prior to the Internet, is an example of technology-driven design. The encyclopedias were more easily searchable and took up less room than the book versions, which had been the norm for many years. However, to access different topics, CDs had to be swapped in and out of the CD drive, and the encyclopedia content could not be updated dynamically when new information became available.

Content-driven design and technology have strengths, but neither one includes the user perspective. Although Taylor stressed the significance of context and the importance of the user, the TEDS framework extended Taylor's initial model by leveraging the concepts of using scenarios and personas from HCI to analyze an information system. These additions integrate a context-specific perspective into the TEDS framework.

Comparing Frameworks

The following sections compare the sub-criteria in the TEDS framework with the attributes of other academic assessment frameworks. This analysis is included to provide insight about the differences between the frameworks, and to show why the TEDS framework was chosen for this study. The decision to use TEDS did not hinge on the fact that it has the most sub-criteria of the evaluation frameworks. A framework can have the most sub-criteria, yet lack important sub-criteria that other frameworks include. In addition, having more sub-criteria requires more time to perform assessments. If there is little difference or unclear differences exist between the sub-criteria definitions, the additional detail becomes cumbersome. It was important to consider how well each framework covered important assessment and evaluation concepts and to compare the level of detail provided in each framework.

Comparison of TEDS and the Assessment Frameworks

The two most notable models of user acceptance frameworks are the Technology Acceptance Model (TAM) and DeLone and the McLean Information IS Success Model (D&M Information Success) . Both born from a systems-level perspective, they were designed to determine the factors that improve the likelihood of information artifact investments succeeding in businesses.

In contrast, the evaluation and assessment frameworks include a greater variety of perspectives, reflecting the diversity of contributing researchers. Concepts from the system-level acceptance models TAM and D&M IS Success appear in some evaluation and assessment frameworks (particularly in the latter ones). The HCI-based frameworks incorporate a more holistic analysis. TEDS provides the most holistic analysis of the frameworks. (See Appendix B for an overview of the user acceptance frameworks and the evaluation and assessment frameworks.)

Comparison of TEDS and User Acceptance Frameworks

Table 6 compares the TEDS framework values with the DeLone and McLean (D&M) IS Success Model. The TEDS ID Map column refers to the TEDS identifier, which maps to the attribute in the corresponding assessment framework. The TEDS framework includes numeric identifiers for the user criteria and sub-criteria. For example, adaptability in the D&M IS Success model corresponds to adaptability in the TEDS framework, which has values in the 400 range. This is noted by the bracket on the left side of the TEDS sub-criteria: “Maps sparsely to Adaptability in D&M.” The D&M model lacks numeric identifiers; thus, the table uses an ‘x’ to identify D&M sub-criteria that maps to the TEDS sub-criteria. Reliability in the D&M model maps to reliability in the TEDS model, which has the value of 304. Usability in the D&M model maps to the ease of use user criteria in the TEDS model, which is comprised of seven sub-criteria (or values).

Table 6. Comparison of TEDS and DeLone and McLean Frameworks

	D&M Inclusion	User Criteria	Value (TEDS)	TEDS ID Map	D&M IS Success Model	
		Ease of Use	101 Browsing/browsability /searchability		**Systems quality**	
		Ease of Use	102 Formatting/presentation	400	Adaptability	
Maps to ***Usability*** in D&M	x	Ease of Use	103 Mediation	303	Availability	
		Ease of Use	104 Orientation	304	Reliability	
		Ease of Use	105 Order/consistency	500	Response time	
		Ease of Use	106 Accessibility	100	Usability	
		Ease of Use	107 Simplicity			
		Noise Reduction	201 Item identification		**Information quality**	
		Noise Reduction	202 Subject description	302	Completeness	
		Noise Reduction	203 Subject summary/ summarization	103	Ease of understanding	
Sparse mapping in D&M		Noise Reduction	204 Linkage/referral	408	Personalization	
	x	Noise Reduction	205 Precision	205	Relevance	
		Noise Reduction	206 Selectivity	503	Security	
		Noise Reduction	207 Order			
		Noise Reduction	208 Novelty		**Service quality**	
	x	Quality	301 Accuracy	301	Assurance	
	x	Quality	302 Comprehensiveness	600	Empathy	
	x	Quality	303 Currency	502	Responsiveness	
	x	Quality	304 Reliability			
		Quality	305 Validity		**Use**	
		Quality	306 Authority		Nature of use	
		Adaptability	401 Contextuality/closeness to problem		Navigation patterns	
		Adaptability	402 Flexibility		Number of site visits	
		Adaptability	403 Simplicity		Number of transactions executed	Use metrics are a proxy for end-user benefits or issues but it can be difficult to determine the correlation between an end-user benefit or issue and a use metric
		Adaptability	404 Transaction			
Maps sparsely to ***Adaptability*** in D&M		Adaptability	405 Trust		**User satisfaction**	
		Adaptability	406 Feedback		Repeat purchases	
		Adaptability	407 Community/social networking		Repeat visits	
	x	Adaptability	408 Individualization		User surveys	
		Adaptability	409 Localization			
		Adaptability	410 Privacy		**Net benefits**	
	x	Performance	501 Cost savings	501	Cost savings	
	x	Performance	502 Time savings		Expanded markets	Net benefits in D&M have a System Stakeholder perspective which is absent from TEDS.
	x	Performance	503 Security		Incremental additional sales	
		Performance	504 Safety		Reduced search costs	
	x	Affection	601 Aesthetics	502	Time savings	
		Affection	602 Entertainment			
Empathy is the only attribute from D&M similar to Affection		Affection	603 Engagement			
		Affection	604 Stimulation			
		Affection	605 Satisfaction/rewarding /incenting			

DeLone and McLean IS Success Model

The D&M framework includes 25 attributes, compared to the 40 attributes in the TEDS framework. Many terms overlap in the frameworks, but there are also noteworthy gaps. In a few cases, attributes from one framework are merged into a term or concept in the other framework. For example, in the D&M model this amalgamation is seen in the term usability, which maps to the ease of use attribute in TEDS. Although the D&M framework omits the detail contained in the seven values of ease of use in TEDS, it does include measures of usage (which TEDS omits), such as the number of site visits or transactions executed.

The significant dissimilarities between the two framework taxonomies are not caused by combining concepts into one term. Instead, the frameworks express the difference between the systems perspective in D&M and the human actor perspective of TEDS. The gaps are evident in both frameworks. Many TEDS values are not included in the D&M framework. Likewise, several attributes in D&M are not included in TEDS.

Most of the D&M criteria in the system quality, information quality, and service quality categories overlap with values in the TEDS framework. However, TEDS emphasizes the user-centered perspective, while D&M takes a more system stakeholder perspective, as seen in several D&M attributes. The D&M net benefits category includes cost savings, expanded markets, incremental additional sales, reduced search costs, and time savings. Although time savings and cost savings can be considered from either the human actor or the system stakeholder perspective, the other criteria benefit the system stakeholder by reducing costs or increasing revenue. This system stakeholder perspective does not appear as cleanly in TEDS.

The D&M model also includes use and user satisfaction categories. Although TEDS includes the affection user criteria, which would seem to map to the D&M user satisfaction category, D&M's

definition of user satisfaction includes repeat purchases, repeat visits, and user surveys. These are system-level metrics of usage, rather than the types of user values found in the TEDS affection user criteria: aesthetics, entertainment, engagement, stimulation, and satisfaction/rewarding /incenting). Seddon's criticism of the D&M model—that use should not be considered a user acceptance antecedent (Rai et al., 2002)—also rings true from the TEDS perspective: Use is not a human actor value, nor does it move an information object further along the information spectrum. In contrast, from the value-added standpoint, system processes do both. Users can be part of the system, which means in some information artifacts, use enables the system to learn and provide better relevance or personalization.

D&M criteria, such as the number of visits, navigational patterns, number of transactions executed, repeat purchases, repeat visits, and user surveys are all metrics that can be used to measure user behavior, but only as a proxy for user engagement, not as intrinsic user values themselves.

D&M's system stakeholder perspective directly maps to the costs and revenues implication for an organization within the framework, the end-user perspective of the TEDS framework, along with its richer sub-criteria, makes it the better option for this study. The TEDS framework addresses organizational implications through constraints that are included in the analysis rather than having explicit system stakeholder sub-criteria.

TAM Model

The TAM model is another seminal user acceptance framework from the literature. The TAM model has a much narrower focus than the TEDS model. All noise reduction, quality, and affection user criteria values are absent from TAM; in sharp contrast, TEDS includes detailed subcategories for user criteria. TAM only sparsely covers the values in adaptability. Table 7 shows a comparison of TEDS and TAM frameworks.

Table 7. Comparison of TEDS and TAM Frameworks

	TAM Inclusion	User Criteria	Value (TEDS)	TEDS ID Map	TAM	
More articulation in TEDS but there is overlap with TAM ***Easy to Use*** in TAM corrsponds to TEDS Ease of Use		Ease of Use	101 Browsing/browsability /searchability		**Usefulness**	
		Ease of Use	102 Formatting/presentation	501	Work More Quickly	Utility definition of ***Usefulness*** is an amalgamation of sub-criteria in TEDS
		Ease of Use	103 Mediation		Job Performance	
		Ease of Use	104 Orientation		Increase Productivity	
	x	Ease of Use	105 Order/consistency		Effectiveness	
		Ease of Use	106 Accessibility		Makes Job Easier	
	x	Ease of Use	107 Simplicity		Useful	
No overlap with TAM		Noise Reduction	201 Item identification		**Ease of Use**	
		Noise Reduction	202 Subject description	105	Easy to Learn	
		Noise Reduction	203 Subject summary/ summarization	400	Controllable	
		Noise Reduction	204 Linkage/referral	107	Clear & Understandable	
		Noise Reduction	205 Precision	402	Flexible	
		Noise Reduction	206 Selectivity	100	Easy to Become Skillful	
		Noise Reduction	207 Order	100	Easy to Use	
		Noise Reduction	208 Novelty			
No overlap with TAM		Quality	301 Accuracy			
		Quality	302 Comprehensiveness			
		Quality	303 Currency			
		Quality	304 Reliability			
		Quality	305 Validity			
		Quality	306 Authority			
Sparse mapping in TAM but TEDS adaptability corresponds to ***controllable*** in TAM		Adaptability	401 Contextuality/closeness to problem			
	x	Adaptability	402 Flexibility			
		Adaptability	403 Simplicity			
		Adaptability	404 Transaction			
		Adaptability	405 Trust			
		Adaptability	406 Feedback			
		Adaptability	407 Community/social networking			
		Adaptability	408 Individualization			
		Adaptability	409 Localization			
		Adaptability	410 Privacy			
		Performance	501 Cost savings			
	x	Performance	502 Time savings			
		Performance	503 Security			
		Performance	504 Safety			
Sparse mapping in TAM		Affection	601 Aesthetics			
		Affection	602 Entertainment			
		Affection	603 Engagement			
		Affection	604 Stimulation			
		Affection	605 Satisfaction/rewarding /incenting			

Obvious gaps exist between the TEDS and TAM frameworks. The TAM usefulness criteria (job performance, increased productivity, and makes job easier and useful) do not appear in TEDS. In combination, these criteria seem to reflect the utility definition of usefulness. In general, the TAM usefulness values are not narrow enough to map to TEDS values. For example, the term useful is as nebulous as the term quality. Without clearer definitions and sub-classification of terms like these, too much leeway exists for the evaluator.

TAM includes fewer user criteria, but does a better job than TEDS of reflecting the utility component of usefulness. Although the aggregated values in the TEDS framework are a measure of usefulness, utility does not emerge as clearly as it does in TAM. The TAM criteria in the usefulness category all point to utility: work more quickly, job performance, increase productivity, effectiveness, makes job easier, and useful. Just as TEDS is an extension to the original Taylor model, the enumeration of user criteria and values is not set in stone in the TEDS framework. As Taylor asserted, over time, more clarity will come (Taylor, 1986) (p. 48). When it does, further framework extensions are expected.

Comparison of TEDS with HCI Evaluation and Assessment Frameworks

Table 8 compares three noteworthy evaluation and assessment frameworks from literature: WebQual (Barnes & Vidgen, 2000), Adaptive Decision Making (ADM) (Hartmann et al., 2007), and Data Quality Dimensions (DQD) Ontology (Wand & Wang, 1996). Table 8 omits columns showing the TEDS framework to make it easier to read. However, the columns with headings “TEDS Map” correspond to the TEDS numeric category identifiers found in Table 6 and Table 7.

These three frameworks—D&M, TAM, and HCI—were chosen for two reasons: they were well represented in the HCI literature, and they all appeared to be good candidates for an iterative evaluation and design study such as this dissertation. All their attributes map to the TEDS framework, but they lack the depth and detail found in the TEDS framework. The increased complexity of TEDS, however, comes at some cost: More attributes mean more definitions and more overhead in using the framework to assess information artifacts.

Table 8. Comparison of TEDS, WebQual, ADM, and DQD Ontology

TEDS Map	WebQual	TEDS Map	ADM	TEDS Map	DQD Ontology
	Ease of use	100	**Usability**	301	**Accuracy and Precision**
101	navigation				
100	ease of use	300	**Content**	300	**Reliability**
	Experience				
600	visual impact	600	**Aesthetics**	500	**Timeliness and Currency**
400	individual impact				
	Information	306	**Reputation/identity**	302	**Completeness**
200	finding information				
300	information content	400	**Customizability**	105	**Consistency**
	Communication and Integration				
204	external integration				
406	communication				

Like the TEDS model authors, the authors of WebQual assert that an assessment framework should delineate how the websites differ in quality. Finer descriptions of which attributes are more or less significant would then be possible, along with knowing where the inflection occurs, rather than only the aggregate rank (Barnes & Vidgen, 2000).

The WebQual authors also state that transaction processing could be included as a criterion for website quality. *Transaction processing* is the capacity of an information artifact to handle a user's need to make a transaction without having to take extra steps or use another site. It measures the number of transactions processed and the number of repeat purchases. This value appears only in the TEDS and WebQual frameworks. It also surfaces as a use metric in the D&M IS Success model. However, TEDS and WebQual include this value from a user-centered perspective rather than from a system perspective.

Use metrics, like the number of transactions, can serve as a proxy for end-user values. It can be difficult, however, to determine the correlation between usage metrics and end-user benefit by merely counting clicks or transactions. For example, an end-user could have more than one transaction on a website because the feature prohibits the purchase of more than one item at a time. The usage metric would suggest something positive because it shows additional usage, but the user may have wanted to purchase more than one item in a transaction.

ADM theory suggests that people's decision making is adaptive and contingent upon the task, context, and their experience (Hartmann et al., 2007). This is also a central assertion in TEDS. Also, like TEDS, ADM uses a Likert rating scale so that participants can rate an information artifact against individual criteria.

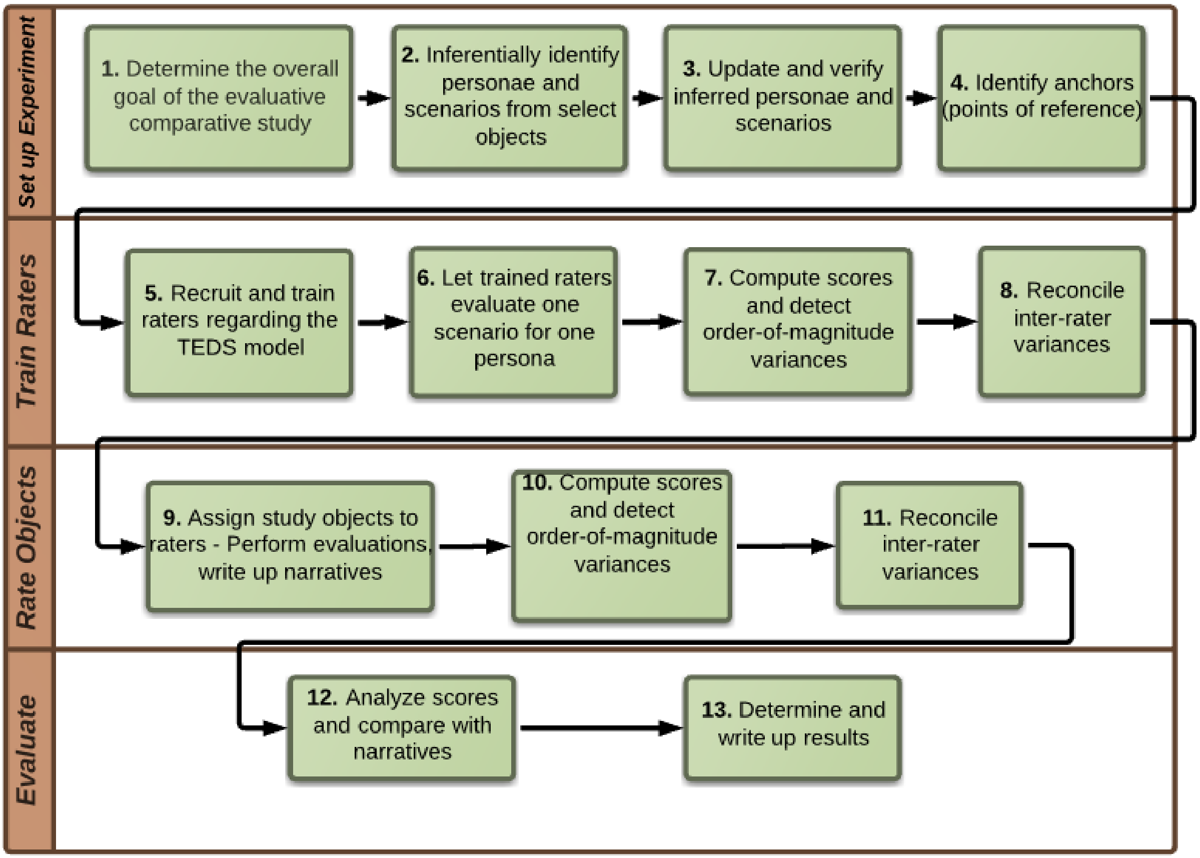
In summary, although there is complete overlap of combined attributes from WebQual, ADM, and DQD Ontology with the sub-criteria in the TEDS framework, these three HCI-based frameworks are missing concepts that were deemed relevant for this study. Therefore, the TEDS framework was deemed superior for this study.

TEDS as an Evaluation and Assessment Framework

Taylor's initial value-added method to assess information artifacts asserted that information artifact assessment is valid when it is grounded in an information use context. The TEDS framework extended Taylor's work by incorporating usage scenarios and personas into the procedure as a proxy for a context-specific perspective.

The TEDS procedure includes 13 steps, which are organized into four parts: set up the experiment, train raters, rate objects, and evaluate results, as shown in Figure 1. The rigor of setup, training, and multiple rounds of inter-rater reconciliation produces expert raters for the assessments. Thus, an expert rater can be defined as someone who has completed this training. Although this expert rating procedure has been used effectively in multiple studies (Scholl & Carlson, 2012) (H. J. Scholl et al., 2014), each study used a relatively small number of raters.

Figure 1. TEDS Procedure Sketch



By following the TEDS procedure, web-based information artifact designers or evaluators can assess a group of information artifacts using four types of analysis:

- General or aggregate observations
- Per-artifact analysis
- Scenario analysis
- Analysis of criteria

TEDS raters are recommended to provide comments and screenshots to explain the reasons for their ratings. This additional detailed information has the potential to be used to generate improvement ideas for the information artifacts. The existing TEDS studies provide insights about which artifacts perform well along the four types of analysis. This study aims to use rater-provided comments and screenshots systematically to make targeted design recommendations.

Some additional issues remain to be addressed with the TEDS method. A gap exists in the TEDS literature regarding which sub-criteria are relevant for assessing an information artifact in a given context. More analysis is required to determine which attributes are more or less significant for a given context. Further, although the TEDS procedure enables pinpoint identification of strengths and weaknesses of an information artifact compared to other information artifacts, a gap exists in the literature that demonstrates how to utilize the output from the assessments to generate feature ideas that can improve the artifact. This, in fact, is a gap in assessment literature in general. Each assessment framework discussed in this literature review can report strengths and weaknesses of an information artifact along some set of criteria in various levels of detail. Yet none demonstrates how their assessment output can be used to generate feature improvements.

In addition, a gap exists in the TEDS procedure itself: Assessments have been implemented only by using small numbers of expert raters. There is an opportunity to compare the assessments produced in this expert method with assessments produced by a larger set of users who more closely match the demographics of target users of the information artifact being assessed. Further, it would be best if a TEDS procedure could gather assessments continuously from users of the information artifact. This approach should be addressed in future research.

Data Mining and Web Analytic Methods for Analyzing and Comparing Information Artifacts

During the past 15 years, many pragmatic approaches to assess information artifacts have emerged from industry. The techniques are relatively new, but discussion is also beginning to appear in the academic literature. These approaches either apply heuristic analysis based on best practices, or they use comparison testing of existing sites to give recommendations. For example, Ivory and Hearst provide a survey of automated usability evaluation methods used in industry (Ivory & Hearst, 2001). The survey includes a taxonomy of tools, which helps inform the practitioner about which tool to use.

Web analytic methods are tools that researchers and development organizations can use to gather real-world information about user behavior on websites. There are multiple ways to use web analytics tools. For example, structured usability testing gathers data by parsing webpages to compare elements with those of another website, or with expert ratings for a page or set of pages in a scenario (Ivory, 2003). Other automated usability tests capture users' mouse clicks, navigation steps, and even eye movements to see how users look at or interact with a website. Automated usability tests have potential as a procedure with the TEDS framework.

Data mining (or transaction log analysis) provides data about a user's interaction with a website. This nonintrusive method for collecting data from many users is now a commonly used tool for websites. "Web tracking collects large amounts of data unobtrusively so that data can be analyzed at a later stage. It can collect data 24/7 without the researcher's presence and without interrupting the information seeker" (Fourie and Bothma, 2007) (p. 270). Data mining is different than other types of evaluation, which ask the user to provide their opinion after completing a task, or someone observes the user and records what occurs.

The advent of web analytics in the software design process has resulted in academic literature (Fourie & Bothma, 2007), (Facca & Lanzi, 2005), (Murphy et al., 2001), (Kohavi et al., 2007), and practitioner books for web developers (Kaushik, 2007), (Eisenberg et al., 2008), (Kaushik, 2010), (Clifton, 2008). Not surprisingly, there is almost no theory in this literature. The literature is mostly how-to resources, which describe best practices for gathering data and recommendations for using the data most effectively. For example, Peral, Maté, and Marco provide an example of a promising technique that uses data mining to identify KPIs in a semi-automated way from existing industry methods that are used to forecast and visualize data correlations (Peral et al., 2016).

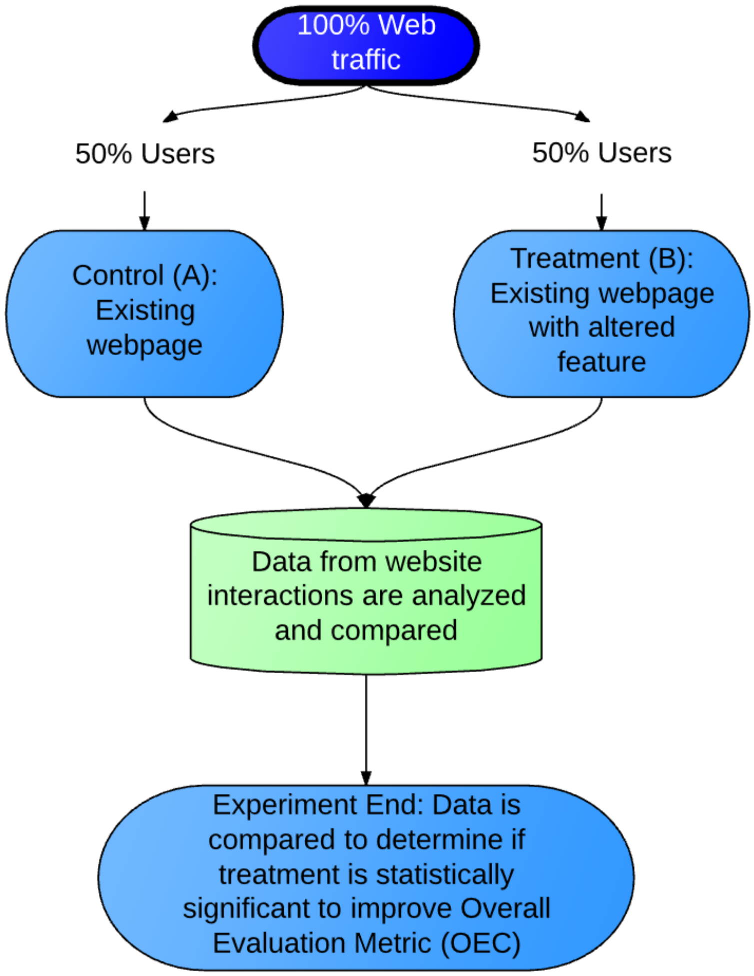
Controlled experimentation—also known as randomized experiments (single factor or factorial design) or A/B tests (Kohavi et al., 2007)—appear to be frequently used procedures that have emerged from this literature. These procedures show much potential as a tool for objectively evaluating the quality of website features. The behavior of users as measured through data mining of predetermined data metrics determines whether a treatment version of a feature performs better than the existing (control) version. “Controlled experiments provide a methodology to reliably evaluate ideas. Unlike other methodologies, such as post-hoc analysis or interrupted time series (quasi experimentation), this experimental design methodology tests for causal relationships” (Kohavi, Henne et al., 2007). When data mining is used in combination with a controlled experiment, different algorithms and designs can be evaluated on a website without participants knowing that they are test participants.

A/B Testing

A/B testing is a data-driven process that decision makers can use to decide whether a feature or feature change should be launched to all users. An A/B test compares two or more versions of an element, such as a webpage or feature, to decide which version performs better. The versions are exposed to the same experimentation simultaneously and measured systematically. The measurements of the two versions are then compared to determine which one performs better for the target audience. When experiments are done using a rigorous A/B or multivariate methodology, researchers can measure the factors precisely by using quantitative methods. They can use experiments iteratively, and once implemented, run them without manual intervention.

For a website, the webpage or feature that is currently live is the control (or the A version) of the experiment. One or more new webpages or features are the B treatment versions of the experiment. Live web traffic is split between the versions in a random fashion to remove bias. Some users see the A version, while other users see the B version. Website metrics are gathered on the control and treatment versions. Statistical tests are then run on the website metrics to evaluate whether statistical significance exists between the variants. This way, the experimenter can accept or reject a null hypothesis that no difference exists between the versions. Figure 2 shows the A/B testing process performed on a webpage.

Figure 2. A/B Testing Process



Many web-based companies use online controlled experiments, including Amazon, eBay, Etsy, Facebook, Google, Groupon, Intuit, LinkedIn, Microsoft, Netflix, Shop Direct, StumbleUpon, Yahoo, and Zynga (Kohavi et al., 2013). For these companies, the question is how they can increase the capacity of their experimentation platform to enable more experimentation, quicker results, and better results.

The body of literature about A/B testing is growing. Researchers at Microsoft and Google have published articles that describe how to efficiently run large numbers of experiments simultaneously on a site (Tang et al., 2010). The literature describes how to do A/B experimentation (McFarland, 2013), (Kohavi et al., 2009), (Crook et al., 2009), (Kaushik, 2006), and also describes the benefits, motivations, and learnings from experimentation (Kohavi & Longbotham, 2011). For example, experimenters at Microsoft provide insight into unexpected results of experiments from Bing and MSN, and explain why A/B tests often show unexpected results that conflict with what the designers and developers expected to happen (Kohavi, 2012). Although experimentation results can show which version of a feature performs better, more discussion about why a feature performs better is needed. More discussion is also needed that helps organizations prioritize which features to experiment with to get the most impact.

Overall Evaluation Criteria

The *Overall Evaluation Criteria (OEC)* is the variable or set of variables that an experimenter will compare between the control and treatment in an A/B test to determine if the treatment outperforms the control. The OEC can be many things, such as the conversion rate, the number of units purchased, or the expected lifetime value. In the case of a TEDS value, the OEC could be a variable or combination of variables that behave as a proxy for user engagement in an experiment. The treatment would be a version of the webpage that demonstrates the value being evaluated.

Whereas user engagement seems to be an appropriate goal for websites and web-based artifacts, sometimes there is an obvious choice for the OEC. For example, if one were evaluating an e-commerce website site, the conversion rate (making a transaction) would likely be a good OEC. The *conversion rate* is defined as the percentage of website visitors who convert from a casual user to performing the desired action on the website (Kaushik 2007) (p. 153).

Data mining can be useful, but on its own it does not offer insight into qualitative aspects, such as user's preferences, affective experiences, and rationale for choices. "In order to really draw benefit from the data collected by transaction logs, we should combine the measurement of the data with an understanding of the rationale behind these measurements" (Fourie and Bothma, 2007). Similarly, A/B experiments provide quantitative metrics but no explanations. Experiments can show which variant is better, but on their own they do not say why something is better.

Although gathering data to track user traffic in aggregate on websites is straightforward, it requires in-depth analysis to determine the meaning from the data (Ivory, 2003). For example, if a user clicks many links on a particular page, does that mean that she is engaged or confused and cannot find what she is looking for? This challenge is discussed significantly in the literature (Kaushik; Kaushik 2007), (Clifton, 2008), (Eisenberg, Quarto-von Tivadar et al., 2008). Kohavi, Henne, and

Summerfield suggest that the long-term effects of an experiment should be considered when determining the OEC. For example, putting ads all over a web page might temporarily increase clicks on the advertisements and increase revenue in the short term. But over a longer period of time, users might abandon the site because they tire of all the ads (Kohavi, Henne et al., 2007). Thus, each experiment requires expert analysis to understand which factors are gathered and measured. The OEC needs to be derived for each variant in an experiment (Kohavi, Henne et al., 2007).

Potential Issues or Pitfalls in A/B Experiment Design

Fluctuations in user behavior exist during the day, on different days of the week, throughout the hours of the day, and even during different seasons of the year (Fourie and Bothma, 2007) (p. 268). A retail e-commerce website has significantly different user behavior during the holidays than it does during the rest of the year. Researchers should factor into the test methodology the duration, time of day, the season, and other factors to ensure that analysis is run long enough to be consistent and not introduce bias into the study. For example, it is legitimate to perform a study during the holiday season if the study aims to determine user behavior specific to that constraint, but if the study seeks to gather more typical user behavior, a time outside of the holiday season is needed to get unbiased data (Fourie and Bothma, 2007).

When examining results, it is essential to be cautious in considering long-term versus short-term effects. Long-term behavior can make a difference on the outcome. For example, heavy ad placement might help the OEC by increasing short-term revenue, but users might abandon the site over time, which would hurt long-term results. Newness and primacy can also be issues. Newness could have users clicking around trying new things in patterns that are not representative of how they will behave over the long term. Primacy suggests that changes to navigation can make users less effective until they learn the new navigation. The Hawthorn effect suggests that if the users are aware that there is a change, they might perceive it in a particular light for a significant period of time, even if there will not be a significant impact longer term (Kohavi, Henne et al., 2007).

A/B testing produces scientific experiments to determine causal relationship between changes and their influence on user observable behavior (Kohavi et al., 2009). A/B experiments focus on one information artifact at a time. The results of an experiment can show if a given derivative treatment of the information artifact performs better than the existing (control) version of the information artifact. Comparing an information artifact against another version of itself is useful, but it would be much better to have information that shows how the information artifact compares with other artifacts that perform in the same context, because it would help shed light on which feature areas are underperforming.

Literature Review Summary

In summary, the evaluation and assessment literature is maturing. Considerable overlap exists between the theoretical frameworks definitions of concepts, such as information quality, usefulness, usability, and ease of use. Each framework discussed in this review, such as the D&M framework, the TAM framework, the HCI frameworks, and the TEDS framework, is theoretically sound and appropriate for assessing information artifacts. Each framework provides a different lens with which to assess information artifacts.

In recent literature, the human-actor and context-specific perspective that emerged from the HCI based frameworks is prevalent. Notable among the frameworks is the TEDS 13-step procedure for evaluating information artifacts. This procedure generates detailed context-specific ratings for information artifacts. Evaluators use a Likert scale from 1-5 to rate how an information artifact performs with regard to a set of criteria for a specific use-case scenario. The ratings allow similar artifacts to be ranked by scenario to determine relative strengths and weaknesses of an artifact. However, there are gaps in the literature regarding how an organization can leverage the evaluation output from the TEDS procedure or other assessment frameworks to generate targeted feature improvement ideas to address weaknesses.

The data mining and web analytics literature is in its infancy. Most of the available literature is how-to books on web analytics, conference papers from A/B testing experts from industry, white papers from organizations about different A/B testing components, and unexpected results or learnings from their experience with A/B testing. However, numerous large software companies use data mining and A/B testing in their development methodology. At the time of this dissertation, there appears to be no A/B testing literature that discusses how an organization can leverage evaluation and assessment frameworks to identify which feature areas should be targeted for A/B testing, nor how to then use the evaluation and assessment frameworks to assist in generating treatments.

Integrating A/B Experimentation into TEDS

Although the TEDS procedure is a pragmatic and proven method for evaluating information artifacts (Scholl & Carlson, 2012), it is manual and time-consuming, and requires significant training to produce experts who can perform the assessments. Web-based information artifacts change quickly. For the TEDS framework to be useful in this dynamic environment, the procedure must match the quick pace of website development and iteration.

Research to extend the TEDS method to leverage the ratings, comments, and screenshots from evaluations to generate targeted feature treatments is needed to make the design process more effective and efficient. For example, an online TEDS assessment tool that could be configured by raters around the world would be beneficial. Such a tool should also allow raters to configure the evaluation to perform a subset of the TEDS categories. This would reduce the time and effort required to perform a full assessment, which leads to the research questions detailed in the next section.

Research Questions

Integration of TEDS with A/B testing is needed to provide qualitative validation of targeted treatment improvements in a trustworthy fashion. Opportunity exists to demonstrate how a practitioner can integrate A/B experimentation into an assessment framework like TEDS to assess the quality of an information artifact iteratively and to identify its competitive weaknesses, and then test potential feature treatments using scientific methods to validate whether the potential change will improve the information artifact's assessed quality. Organizations that are new to software development or already A/B test information artifacts can benefit from procedures that identify areas needing development attention and helps to generate targeted treatments.

Research Question 1 (RQ1): To what extent will the iterative process of "TEDS -> Feature ideation -> A/B testing -> Updating information artifact -> Round 2 of TEDS" produce an improved TEDS assessment of the information artifact?

A gap exists in the literature regarding how a practitioner can leverage the output from an assessment framework like TEDS to generate feature ideas to be created. The TEDS method provides detailed context-specific ratings of information artifacts, along with comments and screenshots from the evaluators to explain how the existing artifact design choices relate to the ratings. What is needed is a process to elicit best-practice targeted design methods for using the ratings, comments, and screenshots from TEDS.

Research Question 2 (RQ2): How can the output from a TEDS analysis be used to generate feature ideas that map to improvements in the TEDS assessment of an information artifact?

A gap exists in the TEDS procedure regarding how to reduce the number of values that are used in the assessment for an information artifact in a given context (scenario, procedure). Previous TEDS studies have used all 40 subcategories when evaluating information artifacts (Jurisch et al., 2014; Scholl & Carlson, 2012; H. J. Scholl et al., 2014). However, these studies noted a desire to reduce the number of sub-criteria used in evaluating each scenario because not all criteria were relevant for each scenario, plus it was tedious to rate 40 characteristics for each scenario.

Research Question 3 (RQ3): How can the number of values used in a TEDS assessment be reduced for evaluating an information artifact in a given context in order to focus on values that are most pertinent to users of the information artifact?

A gap exists in the TEDS procedure regarding how to gather TEDS value ratings data for an information artifact using a method other than expert rating assessments. Reviewers of previous TEDS projects noted potential bias when a relatively small number of expert raters was used to evaluate websites. Investigation of gathering TEDS assessments from a larger number of information artifact users is needed.

Research Question 4 (RQ4): How effective is data from TEDS assessments gathered using surveys from a purposive sampling, in providing effective data to be used in combination with the expert ratings data provided in the TEDS method?

New Literature

Relevant contributions in the academic literature regarding the evaluation and assessment of Information Technology have continued at a steady pace throughout the work of this dissertation, signaling that the research domain remains significant. A discussion of updates to the theoretical contributions is included here in the literature chapter. In chapter 5, a brief review of recent literature is further discussed within the context of the findings.

Theoretical Refinements to TAM, D&M, and Other Assessment Frameworks

Recent literature recommends theoretical changes to existing assessment frameworks and provides examples of applied usage of these frameworks to evaluate information artifacts. In some cases, the theoretical updates appear to be a cross-pollination with learnings from other frameworks' theories. Other motivations are also apparent, as researchers apply the frameworks to different information artifacts in a variety of contexts.

For example, several contributions address today's highly interactive mobile technologies. Samardžija proposed new dimensions and relationships with the D&M Information Systems Success model in her dissertation (Samardžija, 2016). She proposes that the D&M model should include a user experience quality construct as an intermediate success dimension between Information System quality dimension and the Individual benefits dimension. This modification seems warranted from the perspective of HCI and the TEDS framework, since it allows the human actor perspective to be more integrated in the D&M model.

Using the TAM framework as a baseline, Alves, et al. added attributes from other frameworks to arrive at 32 characteristics (Alves et al., 2016). Other studies integrated concepts from different frameworks. Forsgren, et al. used the Systems Administrator context in their study (Forsgren et al., 2016) to further Wixom and Todd's work (Wixom & Todd, 2005)about the integration of User Satisfaction (also known as evaluation and assessment frameworks in this study) and User Acceptance frameworks. Sharma and Lijuan also proposed a theoretical extension to the D&M model for user satisfaction based on analyzing e-commerce websites (Sharma & Lijuan, 2015).

The progress of these other frameworks is pertinent to the future extension of the TEDS model. When developed, the TEDS framework contained a superset of values and characteristics from the other academic frameworks, which still appears to be true. However, it is worth evaluating new values that are advanced in other frameworks. As discussed in chapter 3 of this dissertation, the 40 attributes in TEDS can be overwhelming when evaluating information artifacts. This is one reason this study developed the 12 clusters used in the evaluations. Future research should analyze the mapping of the individual TEDS values and the 12 clusters to the attributes that have been updated in other frameworks, similar to the mappings shown in Table 6, Table 7, and Table 8.

Applied Usage of Theoretical Frameworks

Theoretical frameworks were applied to evaluate information artifacts in several recent contributions. The D&M system was used to evaluate information systems in hospitals (Saghaeiannejad-Isfahani et al., 2015). Researchers in Australia developed a quantitative survey that leveraged D&M and TAM to evaluate eLearning systems (Alsabawy et al., 2016). Hamid, et al. developed a TAM-based questionnaire using a Likert scale to evaluate eGovernment artifacts (Hamid et al., 2016). Rogers used a TAM framework method to test how small business owners in Central Ohio come to accept and use computerized accounting systems (CAS) (Rogers, 2016). Alves, Savaris, and von Wangenheim used their modified TAM framework to evaluate the integrated telemedice and telehealth system in Santa Catarina State, Brazil (Alves et al., 2016). They developed a questionnaire called AdEQUATE that used a Likert scale for user assessments. Their method could leverage TEDS by incorporating scenarios and personas, since it lacks a formal mechanism to incorporate context into the evaluation.

Updates to the TEDS Literature

Two recent studies relate to TEDS directly. Scholl (Margit) et al. integrated the TEDS assessment method in the Moodle learning platform at Technical University of Applied Sciences Wildau (M. Scholl et al., 2014). This is mostly an applied paper. Rather than propose new theory, it shows the framework being used to improve courses in the Moodle learning platform. Nabavi and Jamali made a theoretical contribution to the TEDS model. They recommended that TEDS include three dimensions for adding value by combining basic elements of goals, users, and data and services in a 3D model (Nabavi & Jamali, 2015). Although this extension does not directly impact this dissertation, it should be evaluated in future research as potential additions to the TEDS model.

Chapter 3

METHODOLOGY

3.1. Introduction

The literature review discussed numerous evaluation and assessment frameworks that could have been chosen to research an iterative evaluation and design and A/B testing process. Of the existing frameworks, TEDS was the most comprehensive, with 40 sub-criteria. The fine granularity of the TEDS analysis makes it not only possible to identify strengths and weaknesses in an information artifact, but also to arrive at detailed feature treatments— something that other analytical frameworks with fewer criteria cannot do. In addition, TEDS includes a robust procedure for evaluating information artifacts that has been tested in practice with sports information artifacts.

In spring of 2015, a TEDS assessment project was organized with four graduate students and one professor at the University of Washington, who volunteered to evaluate the 10 iOS mobile applications in this study. There were five phases in this research:

- **Phase 1 - Initial TEDS Assessments:** An assessment of 10 iOS mobile sports applications to produce ratings and rankings across eight scenarios of action/interaction for each application. Included in the set of 10 mobile applications was the Sounders FC iOS mobile application. The Seattle Sounders FC organization is a professional Major

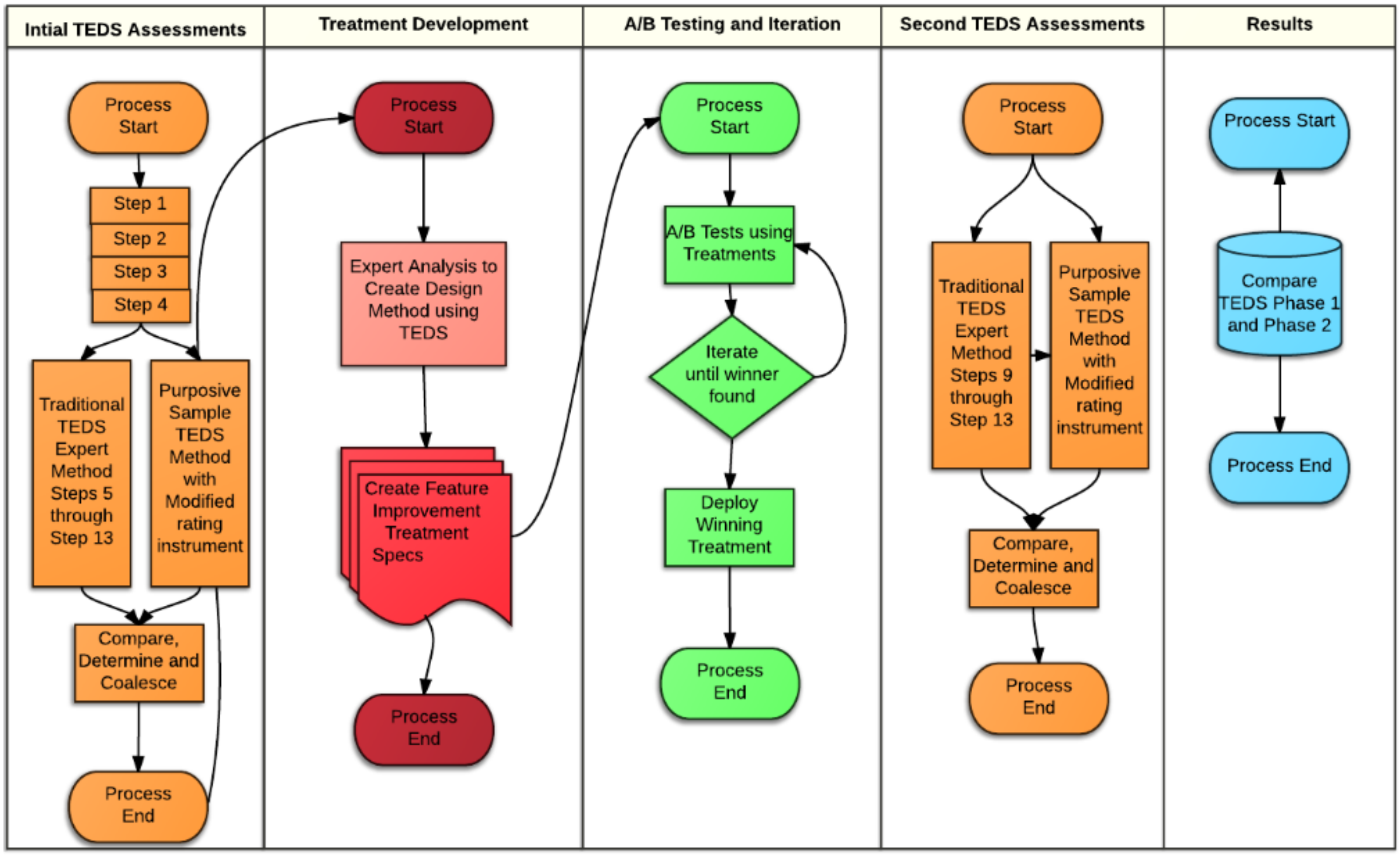
League Soccer (MLS) club based in Seattle Washington. The Sounders FC organization agreed to participate in this study. Assessments were completed in two stages: First, trained evaluators used the TEDS 13-step procedure to evaluate the 10 mobile sports apps. Second, a purposive sampling method was used to recruit users of the Seattle Sounders FC iOS mobile application. The self-selected users completed their ratings using a modified, web-based TEDS survey instrument.

- **Phase 2 - Treatment Development:** A design procedure using the TEDS assessment output to generate feature improvement treatment specifications was developed using a modified Delphi focus group method. The design procedure was developed and communicated such that non-experts could use it. For example, it included a workflow template to help non-expert raters prioritize the screenshots and comments to be used in the identification of design patterns. The procedure also included an example specification that the non-expert raters could copy. The procedure was used to generate feature treatments for the Seattle Sounders FC iOS mobile application.
- **Phase 3 - A/B Testing and Iteration:** During this iterative improvement phase, the Sounders FC organization prioritized the feature treatments that were generated from the treatment development phase, and then selected two treatments to be A/B tested on their iOS mobile application.
- **Phase 4 - Second TEDS Assessments:** A second TEDS assessment phase with the Sounders FC iOS mobile application using the same method from Phase 1 produced new ratings across the Player Information and Schedule, Results, and League scenarios.

- **Phase 5 - Results:** The results from TEDS Phase 1 and Phase 2 were evaluated and compared.

Figure 3 summarizes the process workflow of the five research phases in this study.

Figure 3. The Five Research Phases



TEDS Methodology (Phase 1 and Phase 4)

The TEDS framework consists of four major components:

- The TEDS model of generic human-actor-centric categories and sub-criteria of utilization, which is an update and extension of the original Taylor model

- The 13-step TEDS procedure, which systematically guides the evaluation and comparison of information artifacts
- The case- and persona-specific scenarios of action/interaction, which TEDS leverages from HCI literature (Scholl et al., 2011b)
- The case-specific *personae*, which are narrative constructs that represent specific (groups of) human actors and their needs

The first component of the TEDS model consists of six value criteria and 40 user sub-criteria for designing and evaluating information artifacts, as shown in Table 9 (Scholl et al., 2011a). The six TEDS criteria are: Ease of use, Noise Reduction, Quality, Adaptability, Performance, and Affection.

Table 9. TEDS Value Criteria and Sub-Criteria

TEDS Value Criteria and Sub-Criteria	
Criteria	**Sub-Criteria**
Ease of Use	Browsing/browsability/searchability
	Formatting/presentation
	Mediation
	Orientation
	Order/consistency
	Accessibility
	Simplicity
Noise Reduction	Item identification
	Subject description/classification/controlled vocabulary
	Subject summary/summarization
	Linkage/referral
	Precision/(relevant retrieved) over (retrieved)
	Selectivity
	Order
	Novelty
Quality	Accuracy
	Comprehensiveness
	Currency
	Reliability
	Validity
	Authority
Adaptability	Contextuality/closeness to problem
	Flexibility
	Simplicity
	Transaction
	Trust
	Feedback
	Community/social networking
	Individualization
	Localization
	Privacy
Performance	Cost savings
	Time savings
	Security
	Safety
Affection	Aesthetics
	Entertainment
	Engagement
	Stimulation
	Satisfaction/rewarding/incenting

These criteria and sub-criteria help analyze the characteristics of an information artifact according to how it addresses and corresponds to the needs and wants of supporters, fans, and followers. (See Appendix C for definitions).

RQ3 of this dissertation aims to consider ways to prioritize criteria in each TEDS evaluation. The objective is not to merely eliminate sub-criteria, as that would reduce the fine-granularity of the TEDS framework. Instead, the objective is to prioritize the sub-criteria for each use-case scenario. After each sub-criteria priority is determined, the researcher can focus on the highly relevant sub-criteria and ignore sub-criteria that are not relevant.

Determining the prioritization of the TEDS sub-criteria requires significant effort. Given the scope of RQ1 and RQ2, the committee of the dissertation proposal defense recommended that RQ3 (the investigation of how to prioritize TEDS sub-criteria) should be addressed in future research. However, the expert raters on this project recognized that the TEDS sub-criteria could be clustered for an evaluation, which would reduce the number of individual items to be rated. Each cluster could be re-expanded to expose the finer-grained ratings when more treatment detail was needed. This ability to zoom in and out analytically demonstrates the orderly scalability of the TEDS assessment model.

To cluster the TEDS sub-criteria, the six TEDS criteria was organized in independent card-sorting exercises. This enabled similar sub-criteria to be consolidated in clusters, but remain in their respective criteria category, thus keeping the TEDS meta criteria in place. This exercise reduced the number of rating characteristics for each scenario from 40 sub-criteria to 12 clusters.

The expert raters identified the FC Barcelona and Seattle Seahawk applications as *anchor artifacts*, which are artifacts that provide a baseline for comparison when users evaluate new artifacts. Anchor artifacts are chosen for two reasons: They support all or most of the included scenarios, and they are expected to be average or above average based on basic testing of the application. For these two artifacts, the raters used the full 40 TEDS sub-criteria for each scenario rating, which matched the approach used in previous TEDS evaluation papers. Raters use anchor artifacts to provide perspective about the meaning of a sub-criterion rating for an artifact in a specific scenario. Then, when a rater evaluates a new artifact, they can compare it to the anchor artifact and assign a higher or lower rating based on whether the new artifact performs better for that sub-criterion.

To address the issue that the FC Barcelona and Seattle Seahawks mobile apps were rated using the 40 sub-criteria, while the remaining nine mobile applications were rated using the 12 clusters, the expert raters arrived at the ratings for cluster values for the FC Barcelona and Sounders FC artifacts by calculating the average value of the subcategories that comprise each respective cluster. Table 10 shows the 12 clusters and 40 sub-criteria for the six TEDS criteria. (See Appendix D for definitions.)

Table 10. TEDS Value Clusters and Cluster Criteria

	TEDS Value Clusters and Cluster Criteria	
Criteria	**Cluster**	**Sub-Criteria**
Ease of Use	**Navigation and Findability**	• Browsing/browsability/searchability • Orientation • Mediation • Simplicity
	Structure	• Formatting/presentation • Order/consistency • Accessibility
Noise Reduction	**Identity**	• Item Identification • Subject description/classification/controlled vocabulary • Subject summary/summarization • Precision/(Relevant Retrieved) over (Retrieved) • Selectivity
	Parsimony	• Linkage/referral • Order • Novelty
Quality	**Completeness**	• Accuracy • Comprehensiveness • Currency
	Trustworthiness	• Reliability • Validity • Authority
Adaptability	**Interaction**	• Contextuality/closeness to problem • Transaction • Feedback • Community/social networking
	Customization	• Flexibility • Simplicity • Trust • Individualization • Localization • Privacy
Performance	**Savings**	• Time Savings • Cost Savings
	Confidence	• Security • Safety
Affection	**Attractiveness**	• Aesthetics • Satisfaction/rewarding/incenting
	Enjoyment	• Entertainment • Engagement • Stimulation

Scenarios of Action/Interaction

The third component of the TEDS framework are the *scenarios of action/interaction*, which describe human behavior within a specific context. Scenarios provide an important link between the evaluated information artifacts and case-specific actions and interactions. They capture and identify potential problems, choices, and solutions. For a soccer team website, such scenarios could include, for example, looking up a game schedule or purchasing tickets or merchandise.

The expert raters determined the scenarios of action during steps 2 and 3 of the TEDS procedure. They leveraged previous TEDS studies of sports information artifacts to identify scenarios (Jurisch et al., 2014) (Scholl & Carlson, 2012) (H. J. Scholl et al., 2014). Other scenarios were discovered as the expert raters evaluated potential mobile sports applications to include in the study. In total, the experts identified 11 major and distinct user scenarios of action/interaction for consideration. Seven of these scenarios were chosen from previous studies, and one new scenario (In-Seat Concessions) was included because it was a new feature that the experts believed was a differentiator in this mobile sports application category. The eight scenarios selected for the 10 mobile applications were:

- Team News
- Player Information
- Schedules, Results, and League
- Merchandise or Store
- Ticketing

- Game Day and Stadium Guide
- Social Media Integration
- In-Seat Concessions

Despite differences in business models of professional sport teams in North America and Europe, these user scenarios were found to exist in professional sports teams scanned in both continents (Scholl & Carlson, 2012). This homogeneity provides evidence that they correspond to important scenarios of action/interaction. Only one scenario in the study, In-Seat Concessions, had not been used in a previous TEDS study. It was added because of its novelty and relevance to the in-game experience with a mobile sports application.

The Seattle Sounders FC development team needed to balance their existing backlog of known development work items and bug fixes with the development effort for this project. This reduced both the number of treatments that were selected for experimentation and the pace of developing the treatments (which in turn delayed the A/B testing phase). Thus, due to time constraints, the Sounders selected the Player Information and Schedule, Results, and League scenarios as the two scenarios for the Phase 2 treatment development and Phase 3 A/B testing phases of the project. Their decision to focus only on the two scenarios made it possible to limit the scope of the non-expert TEDS surveys gathered to these two scenarios. The larger selection of information artifacts and scenarios identified in this study from non-expert ratings and TEDS expert ratings will be available for use in a subsequent study, or for the Sounders to use on their own.

Human Actor Characteristics – Defining the Persona

The fourth component of the TEDS framework are the personae. These narrative constructs of typified human actors link the evaluated websites with the needs of *specific human actors*—identifiable groups of people with similar profiles and needs (Scholl et al., 2011a). Personae strive to capture prevalent traits, behaviors, interests, beliefs, and values of specific human actors. The expert TEDS raters in Phase 1 used one persona in their evaluations.

The age of pro sports event spectators is 12-70 years old, with more men (some 60 percent) than women (some 40 percent) (Bernthal & Regan, 2001; Filo, et al., 2009). The demographics of human actors who access pro sports websites appear to be similar, if not identical, to the spectator demographics and categories, except that the age range is narrower based on web literacy and access (Hsu, 2009; Lipsman, 2009). The analytical persona for the traditional expert TEDS procedure was intentionally generic to match the persona used in previous TEDS sports application studies (Scholl & Carlson, 2012; H. J. Scholl et al., 2014).

The fictitious persona using professional sports websites is named Casey, who is described as follows: "An individual between 16 and 60 years of age; male or female; supporter, fan, or follower; average computer literate with unrestricted access to the Internet; interested in at least one professional sports team; accessing pro sports team websites occasionally to frequently; purchasing tickets or merchandise never or occasionally to regularly" (Scholl & Carlson, 2012) (p.

142). The Sounders FC soccer club was provided this description of Casey. Although they know from market research the demographics of their fan base and who attends their games, they did not object to using this generic persona for the evaluations.

TEDS Expert Ratings

Though there have been up to 12 raters in a successful TEDS project (Scholl & Carlson, 2012), the five volunteers for this project was sufficient, given that two previous TEDS studies had produced effective results with five expert raters (Jurisch et al., 2014; H. J. Scholl et al., 2014). The group of five met weekly for 10 weeks in 2015 to implement the 13-step TEDS procedure (Figure 1 in the Literature Review chapter). The first three meetings were two hours long and held at the University of Washington. By the end of the project, the weekly meetings occurred online for an average of one hour.

The first four steps of the procedure involved setting up the experiment. This occurred during the first three weeks that the raters met. Participants discussed and determined the 10 iOS mobile apps to include. They finalized the persona and established the eight scenarios to include. During steps 5-8 of the TEDS procedure, the participants were trained as expert raters. This occurred during the next two weekly meetings. Each participant rated one of the anchor artifacts during the week and provided their evaluations to the lead researcher.

The lead researcher combined the evaluations from each rater in a Microsoft Excel spreadsheet.

The spreadsheet included conditional formatting to highlight ratings that had a discrepancy of more than 1 standard deviation Likert value between raters. These discrepancies were used for in-depth inter-rater discussions of each scenario, which helped raters better understand all components of the TEDS framework, including scenarios, personas, and the TEDS sub-criteria and clusters. After completing the two anchor evaluations, the team developed the TEDS cluster taxonomy that was used in the remaining evaluations.

In steps 9-12 of the TEDS procedure, the raters developed a rhythm for rating the remaining eight artifacts over the next four weeks. Each rater provided an Excel spreadsheet that contained their TEDS ratings, comments, and links to their provided screenshots for each remaining artifact in all eight scenarios. During the weekly meeting, the team discussed the inter-rater reliability for two artifacts. An Excel spreadsheet was generated for each artifact that compiled the ratings, comments, and screenshot references from the individual expert rater spreadsheets. The raters used each consolidated spreadsheet to discuss inter-rater reliability and to generate comparative TEDS ratings. Consistent with the evaluations in previous TEDS projects, the evaluations by the expert raters became more consistent as the weeks progressed, mainly due to the inter-rater discussions helping them develop a more uniform understanding of the context, sub-criteria, and artifact features.

TEDS User Ratings

Reviewers of previous TEDS projects expressed concerns about potential bias when a small number of expert TEDS raters is used. They were also concerned that expert raters do not necessarily match users of the information artifact being evaluated. To address these concerns, a second assessment method was carried out using purposive sampling (referred to as the Facebook Purposive Sample Method) of fans of the Seattle Sounders FC from Facebook and Instagram. A modified TEDS rating instrument was developed that generated scenario-based TEDS surveys. (See TEDSRate survey instrument in Appendix H).

The TEDSRate survey instrument is a web-based tool that runs on any browser. The first page of the survey provides basic information about the study and the survey. If the user chooses to participate, the next page gathers username and password credentials to provide a unique identifier for the user's survey, and to enable the user to return to complete or modify their survey later. Page 3 gathers basic demographic information (level of education, age, gender, and geographic region). Page 4 gathers information regarding how much the user has interacted with the artifact, and how fervently they followed soccer in the past year based on the number of games they attended and how often they watch games.

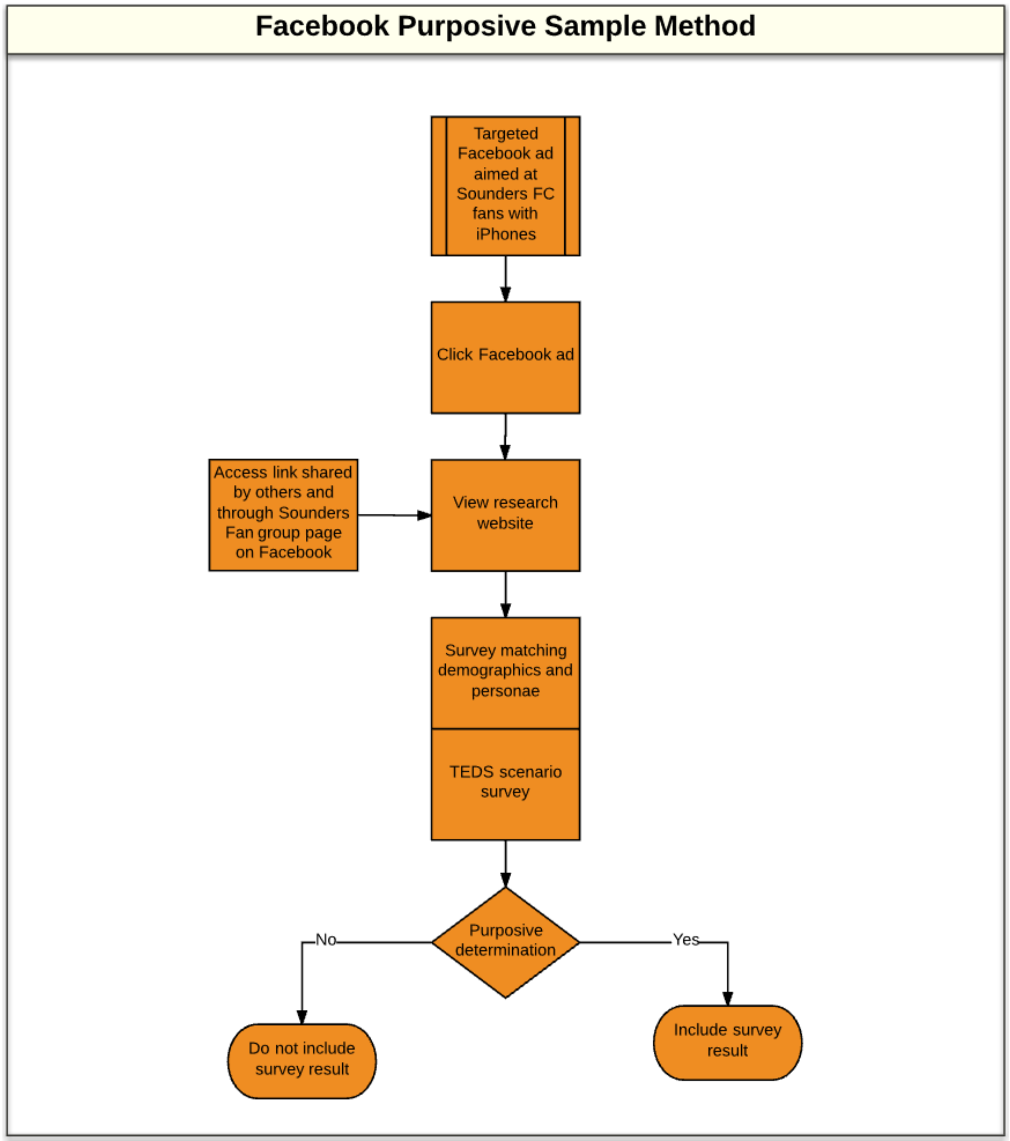
The remaining 12 pages of the survey iterate through the TEDS clusters, gathering the user's ratings, comments, and screenshots for the scenario. Each page includes a layman's definition of the cluster and check boxes to select a Likert value ranging from 1 to 5. The layman's cluster definitions can be seen in Appendix H. Users can use a comment field to explain their rating. They can also use a screenshot control to drag and drop a screenshot picture, or they can browse to select a picture. The development and use of the TEDSRate user survey was performed as a trial of this new assessment method. Using laymen is beneficial because the evaluations are completed by actual users of the artifact. The expert raters, in contrast, performed TEDS evaluations using a persona to represent a context-relevant user perspective. Unlike expert raters, laymen raters were not trained on the TEDS procedure, nor did they participate in inter-rater reconciliation of their ratings to refine their understanding of the components of the TEDS method.

Raters were recruited to the TEDS research website (http://depts.washington.edu/tedsrate/tedsrate/welcome.php) through a post on the Facebook Seattle Sounders FC Fan group page. The TEDS website described the study with links to the Player Information and the Schedule, Results, and League survey instruments. Also, a $5 Amazon credits was offered as an incentive for completing both surveys. This second method of gathering assessments resulted in a larger sample size of ratings with users who self-identified as team fans.

Although investigation of how the user and expert rater evaluations of the Sounders FC mobile application correlate is not the focus of this research, having both rating sources before the treatment and after A/B testing phases made it possible to compare their respective evaluation data for each phase. The survey results collected from this second instrument were compared with the expert-generated analysis and included in the TEDS Phase 1 and Phase 4 expert data sets. Figure

4 summarizes the workflow in this method. Tables comparing the expert ratings to the user ratings for Phase 1 and Phase 4 are found in Tables 20, 21, 26, and 27 in chapter 4 (findings).

Figure 4. Facebook Purposive Sample Method

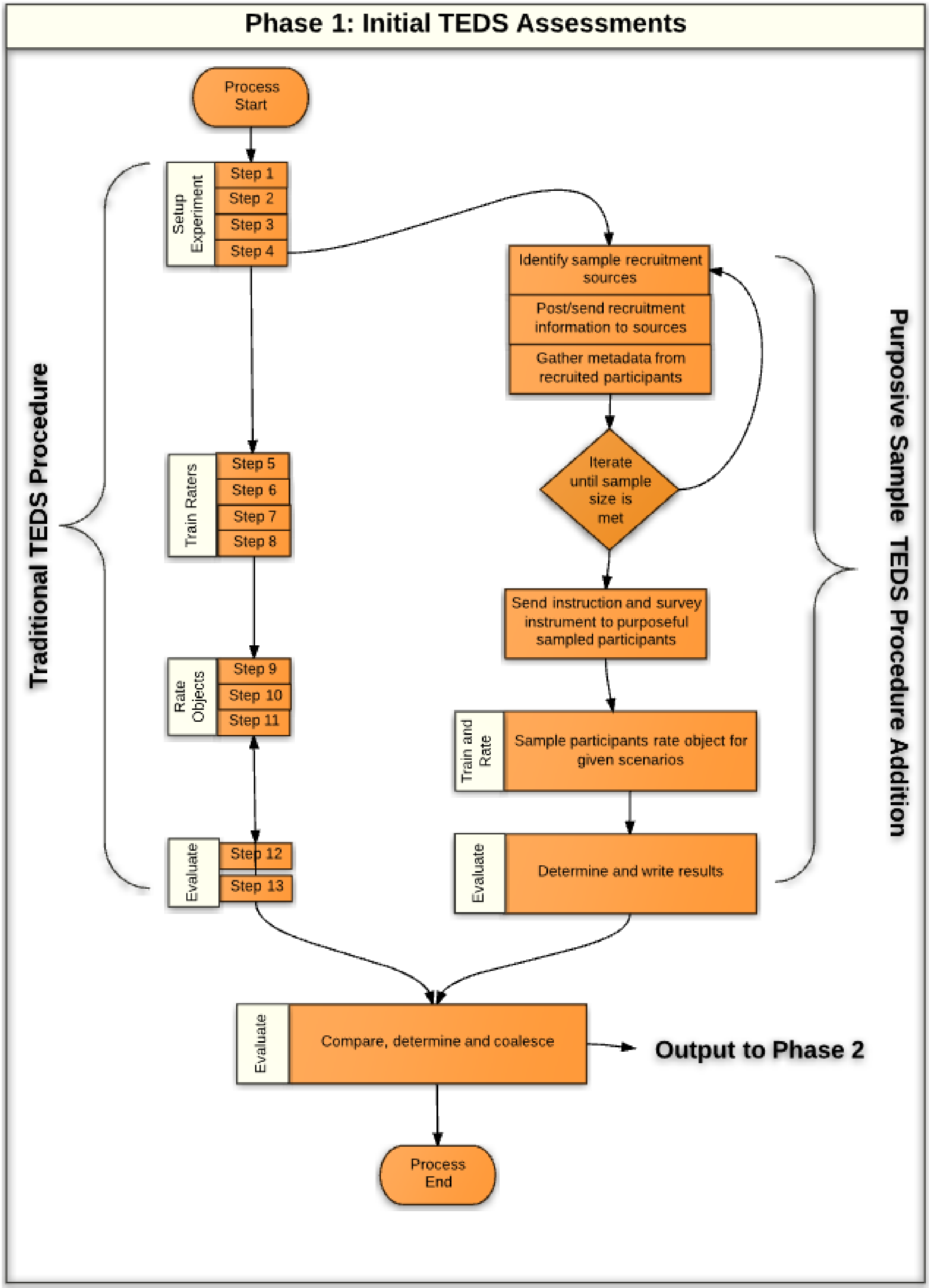


The Facebook ad targeted male and female users ages 16-60 who lived within 50 miles of Seattle and were identified as having an interest in the Sounders FC club and iPhones. In addition, a post on the Sounders FC Fan group on Facebook requested participation in the study from Sounders fans with iOS devices. The post targeted users who identified themselves as Sounders fans by joining the fan group page.

The Sounders have a mobile application for iOS, Android, and Windows phones. The messages posted on Facebook to potential raters stated that the study was for the iOS (Apple) version of the application. No controls were in place to verify that raters used an iOS device, or that they had any version of the application. Since the TEDSRate survey was a browser-based tool and independent of the Sounders FC iOS application, raters self-identified as users of the iOS application. Furthermore, comments on the Facebook page from potential raters indicated that they understood that the study was specific to the iOS version of the application. Numerous users expressed their desire to participate but only had access to an Android or Windows version of the Sounders FC mobile application.

Who to include in the purposive study was based on answers to the seven demographic questions in the survey (see Appendix H), such as how often they watched soccer, and how many soccer tickets they purchased in the last year. Users who rarely or never watched soccer were excluded. The other criteria (they were a fan or follower of a team and had basic computer literacy with access to the Internet) were met implicitly by the Facebook targeting. Given this set of criteria, the purposive sampled users were likely more engaged with the team-specific information artifact than the expert raters were. In addition to the seven demographic questions, one rating question was used for each of the 12 clusters. Surveys submitted with five or fewer cluster ratings were removed. Figure 5 shows the two methods used in Phase 1 for gathering TEDS assessment data.

Figure 5. Phase 1: Initial TEDS Assessment



Treatment Development Phase Methodology (Phase 2)

Phase 2 includes two parts:

1) Produce a design method to generate feature improvement specifications using the TEDS assessments.

2) Use the design method from part 1 of Phase 1 to generate treatment specifications for the Sounders FC iOS application.

Produce a design method to generate feature improvement specifications

The building blocks of this phase include design, the TEDS framework, and A/B testing. Thus, it was a requirement to acquire expertise for each these 3 domains. Because this method was to be applied to the Sounders FC iOS application, it was also determined that mobile application development expertise, as well as representation from the Sounders FC organization was needed to collect perspective from business stakeholders in the process.

The Delphi method was chosen as the research method for this phase. Delphi is a widely used and accepted method for gathering data from users with domain expertise (Hsu & Sandford, 2007). There is some criticism that the approach does not always optimize for the best idea because, in its traditional form, it is designed to reach consensus among the participants, which can lead to compromise (Sackman, 1974). However, this study needed a method to bring unique domain knowledge together using an efficient brainstorming framework. The emphasis was to bring together experts in TEDS, A/B testing, and mobile application development in a collaborative discussion, so that the unique requirements from those domains were included. Compromise because of consensus building was unlikely to be an issue. In traditional Delphi studies, participants' ideas are ranked in multiple rounds of anonymous surveys. The most popular ideas remain for the next round. However, the use of Delphi in this study did not rank ideas. There was only one round of surveys, which expedited gathering information from participants. The gathered survey results were provided to the participants at the focus group brainstorming session. The Delphi method was suited to efficiently structure this merging of domain expertise.

The Delphi method was developed in the 1950s by Norman Dalkey of RAND Corporation, who was tasked by the U.S. military to gather expert opinions from Soviet strategic planners about the quantity of A-bombs they would need to produce in relation to one aspect of the cold war conflict with the US. The Delphi method gathers anonymous input from experts in an iterative process of data collection, analysis, and feedback to gain insight or knowledge about a problem or phenomenon (Skulmoski et al., 2007). The four components of the classical Delphi method include:

- **Anonymity of Delphi participants:** Participants can express their opinions freely without the influence of others in a group setting.
- **Iteration:** Participants have opportunity to refine and adjust their own responses based on the input of others.
- **Controlled feedback:** Participants get feedback from others in a controlled manner to reflect and clarify their own views.

- **Statistical aggregation of group response:** Quantitative analysis and interpretation can be performed with the data gathered from the experts.

Although Classical Delphi projects adhere strictly to these four characteristics, many studies have successfully leveraged parts of the Delphi method while varying from strict compliance to only these four characteristics (Skulmoski et al., 2007). The basic principles of Delphi provide a structure for group communication to enable individuals as a whole to deal with a complex situation (Okoli & Pawlowski, 2004). In their literature review, Okoli and Pawlowski provide multiple examples in which the Delphi method was used effectively to develop an information artifact concept or framework.

Rather than rely on a statistical sample to represent any population, the Delphi method relies on experts who have a deeper understanding of the topic being investigated. Expert panel members are chosen based on criteria that is central to the research question. After a list of viable participants is identified, they are invited to participate. The number of recommended participants included in a Delphi study varies from six to 18 members.

An instrument is developed to gather answers to the open questions of the study. In a classical Delphi study, interaction with the participants is done anonymously and iteratively through multiple phases. Schmidt’s (Schmidt, 1997) process for implementing the Delphi method divides the iterations into three phases, as shown in Table 14 (Okoli & Pawlowski, 2004).

Table 14. The Three Phases of the Delphi Method

Phase 1: Brainstorming	▪ Treat the experts as individuals, not panels (for Phase 1 only) ▪ Provide the questionnaire to all experts for their opinion ▪ When questionnaire data is returned, consolidate, remove duplicates, and unify terminology ▪ Resend refined data to the individual experts to verify that it was interpreted correctly
Phase 2: Narrowing down	▪ Have the experts work with their panel teams ▪ Refine the data by expert group to solidify and rationalize the data from the expert perspective
Phase 3: Ranking	▪ Have all team members participate together to reach consensus and coalesce interpretations ▪ Refine the data as a group ▪ Use a statistical tool (if needed) to satisfice agreement

The classical Delphi method includes a prescriptive set of steps to ensure anonymity and a method to produce prioritized or ranked lists from the experts participating. However, for this study, the emphasis was to bring together specific expertise to brainstorm a design method. A modified Delphi method using design expert participants and an initial questionnaire instrument followed by an in-person focus group were chosen to provide structure to group communication, yet allow for expedient discussion among the panelists. The goal was to glean learnings from the phase about how to use the assessments effectively to generate feature improvement ideas.

- **Step 1: Identify and recruit Delphi panel members**

A request for a human subjects exemption application was sent to the Human Subjects Division at the University of Washington for the use of the Delphi focus group in the treatment development phase. This research fit into Exempt Category 7 for non-federally funded research that includes the evaluation of data and brainstorming. The exemption was granted on January 28, 2016.

The aim of this phase was to facilitate a brainstorming session with participants who had expertise that was relevant to the major design components of this study. Thus, to qualify for inclusion in the Delphi panel, participants had to meet at least one of the following five criteria:

- Software design expertise
- TEDS expertise
- Mobile application development expertise
- A/B testing expertise
- Representation from the organization participating in the A/B testing phase

This study did not require a large group of participants, provided all five criteria had at least one expert. The panel needed 6-10 participants for an interactive focus group discussion. Candidates were recruited using convenience sampling—a type of non-probability sampling technique that is based on the judgement of the researcher (Coyne, 1997). The presence of Microsoft, Google, Amazon, and other high tech companies in the region provided access to many experienced designers and A/B testing experts.

Participant recruitment began the first week of January 2016, when six qualified participants were contacted by email to determine their availability and willingness to participate. These candidates, whom the researcher knew from industry and academic work experience, were also asked to recommend other qualified candidates who might be interested in participating. This snowball recruiting produced six more qualified candidates. A study description and request for participation were posted on distribution lists. One candidate from Google posted the solicitation to distribution lists at Google for Seattle-area candidates with design and/or AB testing expertise. Candidate recruitment stopped after the five criteria were met and there were at least six participants.

Through these processes, a Delphi focus group workshop was created that consisted of seven information artifact design experts and assessment experts and one iOS developer from the Seattle Sounders FC organization. The participants included:

- A software design expert from Microsoft with 20 years of experience working on desktop and web application design.
- An industrial design expert from the University of Washington who is a PhD candidate at the Information School (iSchool).
- An A/B testing expert from the Microsoft Bing A/B Experiment Platform team with four years of A/B experience and 10 years of industry experience in software design.
- The lead iOS Mobile application developer from Sounders FC with more than 20 years of software development experience.

- Two TEDS researchers from the University of Washington iSchool who were in their second year of the Masters of Science in Information Management program. They were also expert raters who participated in the TEDS assessments for this study.
- Two experts in the TEDS methodology from the University of Washington iSchool. One of the experts was a professor at the University of Washington; the other was the researcher of this dissertation.

- **Step 2: Introduce TEDS evaluation and assessment framework to Delphi panel members**

Training was developed on the TEDS evaluation and assessment framework for Delphi panel members who were unfamiliar with TEDS. (See the Training PowerPoint deck in Appendix P.) It was not necessary for panel members to be experts in rating artifacts with TEDS, but they needed to understand concepts used by TEDS, such as scenarios and the cluster criteria the mobile applications were being rated on, and where to find cluster criteria definitions. They also needed to understand the output format that is generated by a TEDS assessment because they would use this output in the design model to generate feature treatments. Each participant met with the researcher for 30 minutes to cover the TEDS training materials. Participants were shown the TEDS ratings, comments, and screenshots posted by the TEDS raters, and a tool for searching and filtering the comments and screenshots by application, scenario, and cluster criteria.

- **Step 3: Iterate with Delphi panel members**

One week after the training and one week prior to the scheduled focus group meeting, the Phase 1 questionnaire (Appendix E), which contained open-ended questions, was emailed to all Delphi participants, with the request that they return their completed questionnaire within three days. This provided adequate time for the results to be compiled and analyzed prior to the focus group discussion, yet kept the activity recent enough to be fresh in their minds during the focus group. All questionnaires were returned within three days. The researcher accumulated, collated, and tagged all responses in an Excel file (Appendix F) and printed paper copies of the anonymized results for each participant to use during the focus group discussion. Table 15 shows the three phases of the Delphi focus group method used in this study.

Table 15. Three Phases of the Delphi Focus Group

Phase 1: Brainstorming ▪ Questionnaire	For this phase, the panel experts were treated as individuals, not as a panel. Each participant filled out the questionnaire, and emailed it to the researcher prior to the focus group meeting: *Q1* ▪ *What method do you use for generating design ideas?* ▪ *Is it a phased method?* *Q2* ▪ *What modifications, if any, do you make if the design is modifying an existing application versus a new application?* *Q3* ▪ *If you have done A/B testing, what are unique things to consider when designing treatment versions for an experiment?* *Q4*

	▪ *Given what you know about the TEDS evaluation data and the goal of this focus group to generate method for creating feature improvement specifications, what steps would you use?* Responses from the returned questionnaires were anonymized in an Excel file. Similar themes were tagged to make it easier to group and filter the responses during the focus group discussion.
Phase 2: Narrowing down ▪ Focus group discussion	Focus Group meeting: All team members worked together as a panel. Each received consolidated comments from the questionnaire. After reviewing comments and refinements together, they discussed the feedback and suggestions.
Phase 3: Ranking ▪ Procedure created ▪ Feature improvement spec prototyped using procedure	Focus Group meeting: The focus group brainstormed a method to refine and reach consensus on the step-by-step procedure for generating feature improvement specifications. The focus group then used the procedure to create an outline for a feature improvement specification for the Seattle Sounders iOS application.

The Phase 1 surveys were completed independently by the participants prior to the focus group meeting. The researcher anonymized the results and tagged themes from the results to make it easier to group and filter responses for the focus group discussion. Phases 2 and 3 occurred during the 2.5-hour focus group meeting, which took place February 5, 2016. The focus group meeting was an iterative process that used brainstorming, focus group discussion, and ranking of ideas. The meeting was organized into three segments:

- **First segment:** Participants reviewed and discussed the responses to the questionnaire from the analyzed compiled results in Appendix F.
- **Second segment:** Participants brainstormed about a method to create treatment improvement specifications.

- **Third segment:** Participants discussed how they would apply this method to the Sounders FC mobile application.

The method to generate treatments is described in chapter 4 (findings), since it was an output of the Delphi method study.

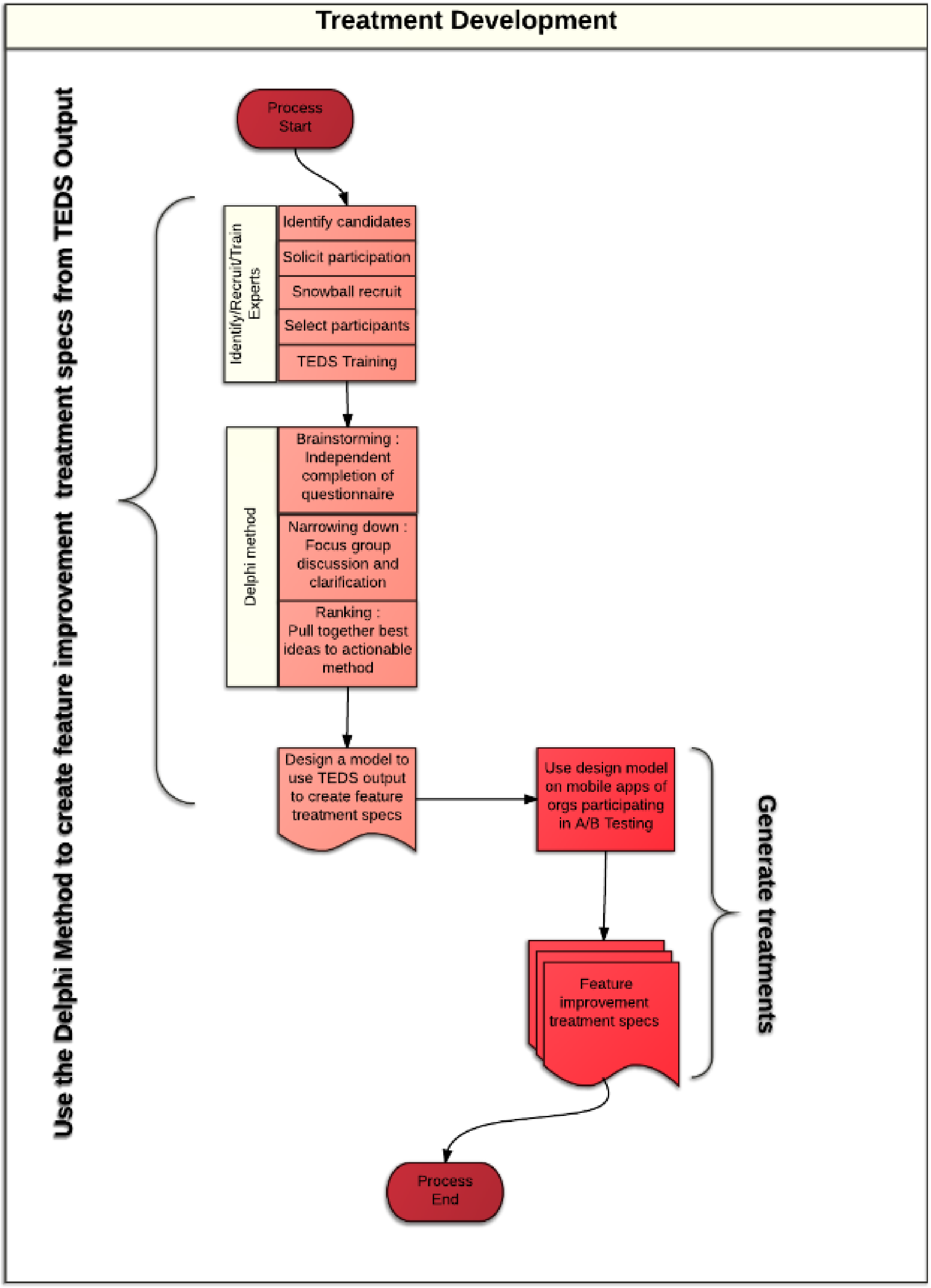
At the focus group session, a whiteboard and a large poster board were used to gather and organize concepts from the participants. After the focus group session, photos were taken of the whiteboard and poster board. The focus group session was digitally recorded. In addition, two graduate students from the iSchool took notes of the session.

Generate treatment specifications for the Sounders application using method.

Following the focus group meeting, the researcher used the feature improvement specification outline and the step-by-step procedure generated in the focus group meeting to create two scenario-based feature improvement specifications for the Sounders FC iOS application: Player Information and Schedule, Results, and League. As mentioned previously, these are detailed in chapter 4 (findings) because they were outputs of the developed method.

Upon completion of Phases 1-3 of the Delphi method, the researcher met with the Sounders FC development organization to present the assessment results and to propose the feature treatments for A/B testing. Figure 6 graphically represents the treatment development phase.

Figure 6. Phase 2: Treatment Development



A/B Testing Phase Methodology (Phase 3)

It is the intention of this project to leverage familiar and respected tools that other researchers can easily replicate. A/B testing platforms range from rudimentary to very sophisticated. Table 16 shows several worthy platforms that could have been chosen for this phase. These are listed because they are market leaders (Optimizely at 39.3%, Adobe Target 20%, Visual Website Optimizer 9.9%, Monetate 4.7%) (Chaffey, 2016) and are based on use experience from the researcher. There are certainly other worthy A/B testing platforms that are not included in the table.

Table 16. A/B Testing Platforms

A/B Test Platform	URL
Adobe Target	http://www.adobe.com/marketing-cloud/testing-targeting.html
Google Analytics Content Experiments Framework	https://support.google.com/analytics/answer/1745147?hl=en&topic=1745207
Optimizely	https://www.optimizely.com/
Mixpanel	https://mixpanel.com/
Monetate	http://www.monetate.com/
Visual Website Optimizer	https://vwo.com/
Yahoo Flurry Analytics	https://developer.yahoo.com/flurry/

The choice of web analytics tool depended on the development stack used for the iOS version of the Seattle Sounders FC mobile app and the priorities of the Seattle Sounders FC development team. Google Analytics Content Experiments was suggested in the dissertation proposal because it is free, commonly used, and well-respected. It allows the analysis of up to 10 versions of a landing page at one time. In an A/B test, the variable or set of variables (the Overall Evaluation Criteria, or OEC) is determined for the landing page, and then the control and treatment are tested and compared. The OEC can test which version of a landing page causes the greatest improvement in conversions or in a specified metric.

The Google Analytics Content Experiment tool provides a visually rich user interface (UI) with graphs and statistics that show the progress of the experiment and features that indicate when the experiment has reached completion based on the confidence level provided by the researcher. Google Analytics Content Experiments platform is a recommended A/B testing platform for projects that use this dissertation as a template for an iterative evaluative, explorative, and generative design process with A/B testing.

However, for this project, the Sounders FC development organization chose the Yahoo Flurry Mobile Developer Suite to perform A/B testing. The A/B platform decision was up to the Sounders FC development team. They were already using Flurry mobile analytics for their mobile applications; thus, it required little additional development work to perform A/B testing. Although at the time of this study, Flurry did not include built-in A/B testing features in its web-based client, it did provide the ability to create segments and then randomly assign users to the segments through an unbiased method. Users in these segments could then be shown treatments or the control through events that Flurry generated. Flurry can track rudimentary user-level and page-level metrics in these segments for A/B testing analysis.

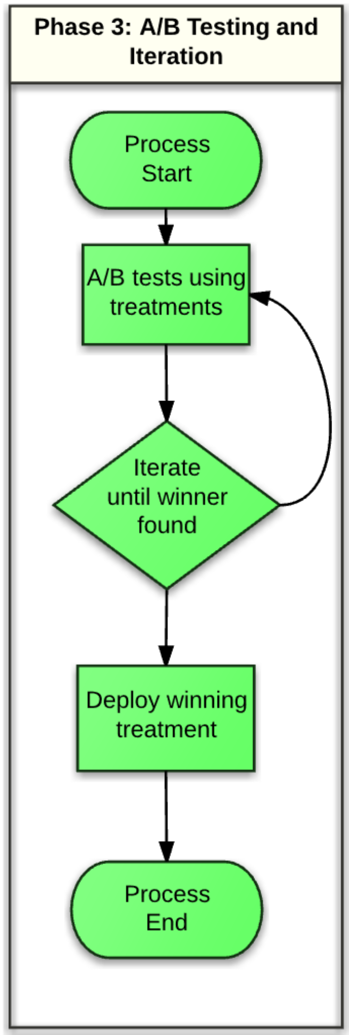
The implication for this project is that the A/B analysis was done manually rather than in an automated fashion. More sophisticated A/B platforms provide more built-in metrics for the OEC. Nonetheless, the study assumed that Flurry mobile analytics is a reliable aggregator of the data gathered for A/B testing.

For each A/B experiment, one or more new versions of the application page were created and hosted on a test site. Each experiment required the following:

- Determination of the percentage of users included in the experiment
- Determination of the OEC
- Assignment of users included in the A/B experiment
- Comparison of the treatment and control performance with the OEC

Figure 7 shows the steps in the A/B Testing and Iteration phase.

Figure 7. Phase 3: A/B Testing and Iteration



Two main factors limited the number of individual A/B experiments that could be run in this study. First, the Sounders FC development organization had existing business priorities, which needed to be completed before they would dedicate time to the A/B tests. This pushed the date of the A/B experiments late into the MLS soccer season. Second, the A/B tests needed to be completed before the end of the season. After the season was over, use of the Sounders FC mobile application would change significantly, so any A/B experiments would have to be postponed until the next MLS season started four months later. One month (October) was allocated for A/B experiments at the end of the Sounders FC season.

The treatments selected by the Sounders for A/B testing were significant in scope. For example, in one experiment, a page on the mobile application presented information in a very different orientation and included more than one way to interact with the page. Bundling multiple changes into one treatment made it impossible to delineate the specific impact of the individual changes. It is possible that specific changes within the bundle of changes were not positive at all. Each experiment could only determine if the combined treatments improved the application. However, in the case of each of the two treatments proposed, the combination of the changes was intentional by the Sounders FC development team. Future experimentation could be performed to isolate individual treatments to test their impact.

Determining the Percentage of Assigned Users to Treatment and Control

All users of the Sounders FC iOS mobile app were included in the experiments, with 50 percent assigned to the control and 50 percent assigned to the treatment. The users either received the treatment for both experiments, or they received the control for both experiments. A/B testing does not require equal percentages to be assigned to the treatment and control, but increasing the number of participants shortens the required duration of the experiment.

Users were assigned odd and even values from the unique device identifier. Assignment to the treatment or control remained persistent throughout the experiment so that if a user returned to the feature on the device later while the experiment is running, they remain with the same assignment. One limitation of this assignment method is that users can become biased in subsequent experiments. To eliminate this bias, it is recommended that user assignments are uniquely generated for each experiment. It is also recommended that the Sounders FC organization adopt a more full-fledged A/B experimentation platform that provides more flexibility in assigning users in experiments.

Determining an OEC for the A/B Experiments

On a transactional webpage, the goal is often to get the user to click a button to purchase something. The OEC can be tied to clicking a submit button or a link to the shopping cart to make a purchase. This makes it easy to determine the OEC. However, for games and entertainment applications like the Sounders FC mobile application, the aim is to engage users, which is not as straightforward. Thus, the challenge is to find an OEC that approximates user engagement and delights with the feature. The number of sessions, the average session length, and the number of page views were considered candidate metrics for the OEC for the Sounders FC application.

The number of sessions over time could be a reasonable OEC if one assumes that the user continues to use the features over time if their needs are being met. Session time can be a problematic measure because the user might find the information they are looking for quicker on a better designed page. The number of page views could also be a reasonable OEC if one assumes that the user is more engaged with the feature if they hit more pages in the application with the new treatment. The concern, however, is that users who are struggling to find information could be clicking many pages because of a poor design or layout. Return users over an extended period could also be a measure of engagement and delight. This measurement requires that the data analytics tool can track return users over time, which, at the time of this study, Flurry Analytics could not do, or at least it was unclear how to track users over time based on how the Sounders FC development organization was using Flurry.

Both A/B experiments in this study were related to the Player Information scenario. The major features in this scenario include the roster of players and the individual player cards. After discussion with the Sounders FC development team, it was determined that the OEC for both experiments would be the average number of player cards selected by each user included in the experiment. Session time was determined to be inconclusive, since it would not be clear if a user spent more or less time on a well-designed player card. However, it was assumed that the average number of player cards selected by a user over time is higher if the user is engaged and finding relevant information on the player card. Conversely, if the user becomes frustrated with the player card because it takes too long to find the information they seek, they hesitate to continue using the feature. This situation demonstrates how the average session time can be misleading. In contrast, the average number of player cards selected would correctly correlate with less user engagement.

The Sounders FC iOS mobile application was built using the XCode and storyboards. In this stack, the application pages are dynamically generated at runtime. By default, there are no page metrics for an analytics program like Google Analytics to add tracking script in the header of a static page without additional development effort. Instead, the development team approximated page-level metrics by adding events to specific objects in the application to simulate these metrics. These events can then be tracked by Flurry. The *Player Detail – Visible* event would occur when the Sounders iOS application generated the player card. Thus, this event was used as the OEC in both A/B experiments.

In summary, there were five phases to this research:

- **Phase 1 - Initial TEDS Assessments:** An assessment of 10 iOS mobile sports applications to produce ratings and rankings across eight scenarios of action/interaction for each website. Assessments were completed in two stages: First, trained evaluators used the TEDS 13-step procedure to evaluate the 10 mobile sports applications. Second, a purposive sampling method was used to recruit users of the Seattle Sounders FC iOS mobile application. The self-selected users completed their ratings using a modified, web-based TEDS survey instrument. The TEDS expert assessment data from Phase 1 was used as input for the Phase 2 Treatment development.
- **Phase 2 - Treatment Development:** A design procedure using the TEDS assessment output to generate feature improvement treatment specifications was developed using a modified Delphi focus group method. The design procedure was designed and

communicated such that non-experts can use it. The developed design procedure was then used to generate feature treatments for the Seattle Sounders FC iOS mobile application.

- **Phase 3 - A/B Testing and Iteration:** During this iterative improvement phase, the feature treatments generated from the treatment development phase were A/B tested on the Sounders FC iOS mobile application.
- **Phase 4 - Second TEDS Assessments:** A second TEDS assessment phase of the mobile applications was used to produce new ratings and ranking across the Player Information and Schedule, Results, and League scenarios for the Sounders FC mobile application.
- **Phase 5 - Results:** The results from TEDS Phase 1 and Phase 2 were evaluated and compared.

Chapter 4

FINDINGS

The findings from the four phases were used in response to the research questions posed in chapter 2.

Addressing Research Question 1 (RQ2)

To what extent will the iterative process of "TEDS -> Feature ideation -> A/B testing -> Updating information artifact -> Round 2 of TEDS" produce an improved TEDS assessment of the information artifact?

Phase 1: TEDS Assessment

The Phase 1 TEDS assessments took place in spring of 2015. The 10 iOS mobile applications that were evaluated included six professional soccer teams, two prominent soccer sports websites, one professional basketball team, and one professional football team. They rated the English version of the iOS mobile applications. For each mobile application, their evaluations produced detailed ratings for a set of eight user scenarios. The combined ratings provided rankings of the mobile applications in each scenario. The assessment output included comments and screenshots from the raters to substantiate and explain their ratings.

Meanwhile, the Sounders FC iOS application continued to evolve during the project, with the deployment of new features from their existing development backlog. Some of those features were influenced by the TEDS evaluations, even though the development team did not apply the A/B testing phase to those features. The implication is that the second round of TEDS assessments in Phase 4, and the subsequent presentation of data and discussion, should include the modified non-A/B tested features, along with any A/B tested features that were generated from this study.

TEDS Expert Ratings

An IT artifact's TEDS rating for a given scenario can be found by calculating the average of its 12 cluster Likert ratings. Table 17 shows the ranking of the information artifacts by the Phase 1 expert raters, who evaluated each scenario on a Likert scale of 1 to 5 (with 1 being low or poor value and 5 being high or excellent value).

Table 17. Phase 1: TEDS Average Expert Ratings by Scenario

Phase 1: TEDS Average Expert Ratings by Scenario												
	iOS Application											
Scenario	**Sounders FC**	**Seahawks**	**FC Barcelona**	**Real Madrid**	**Chicago Fire**	**FC Bayern Munich**	**Onefootball**	**ESPN FC**	**Brooklyn Nets**	**Arsenal**	**Average rating**	**Std dev**
Player Information	3.35	3.45	3.58	**3.79**	2.79	3.04	3.64	**2.17**	3.36	3.13	3.23	0.48
TeamNews	3.36	**3.69**	**3.64**	3.30	**2.81**	**2.92**	**3.75**	3.13	3.30	3.03	3.29	0.33
Schedule, Results and League	3.25	3.83	**4.10**	3.38	**2.89**	**2.98**	**3.95**	3.19	3.72	3.25	3.45	0.42
Merchandise and Store	3.20	**3.74**	**3.61**	3.14	**2.94**	3.14			3.09	3.09	3.24	0.28
Ticketing		2.66	**4.23**	3.53	2.78				3.35	2.73	3.21	0.61
Social Media Integration	**2.74**	2.98	**3.44**	3.33	3.18	**2.66**	**3.39**	**2.63**	3.17	3.00	3.05	0.30
Game Day and Stadium Guide		3.19			**2.58**	2.66			**3.48**		2.38	0.43
In-Seat Concessions		3.15			**2.06**				3.52		2.91	0.76
Average Likert	3.18	3.34	**3.77**	3.41	**2.75**	2.90	**3.68**	**2.78**	3.37	3.04	3.22	0.36

One can compare the average Likert rating of each IT artifact by scenario (row). For example, Real Madrid had the highest Player Information rating among the 10 iOS applications, while ESPC FC had the lowest rating. The number in each cell shows the average rating of all clusters for the iOS app in that scenario. Darker green cells indicate a higher (better) rating for the column team's artifact. Lighter yellow cells indicate lower ratings. Numbers that are black and bold are at least 1 standard deviation above the average rating for all mobile applications for that scenario. Numbers that are red and bold are at least 1 standard deviation below the average rating for all mobile apps for that scenario. Grey cells indicate that the mobile application does not support the scenario.

The Onefootball mobile app, for example, supports only four of the eight scenarios, yet it rates highly for the four scenarios that it supports. The total averages in the table do not include ratings for non-supported scenarios. Thus, the average Likert score remains high for the Onefootball mobile app. The full TEDS analysis includes tables that include 0s for the non-supported scenarios. However, for this study, with the goal of generating treatment ideas in the subsequent phase, it was more useful to compare only the artifacts that support the scenarios. The Sounders FC iOS mobile application was in Beta form during the Phase 1 expert ratings, so the Ticketing scenario and Game Day and Stadium Guide scenario were not available for rating.

The Sounders FC chose Player Information and Schedule, Results and League as the highest priority scenarios for Phase 2 treatment development and Phase 3 A/B testing. Users interacted frequently with these features, as evidenced by Flurry event logs. The other six scenarios were considered out of scope for those two phases of the project. However, all eight scenarios are included in the TEDS expert evaluations performed during Phases 1 and 4.

The average expert Likert rating for Player Information was 3.23. The Sounders FC mobile app was just above this average, with 3.35. Although this is not a terrible rating, it was in the five lowest-rated applications for this scenario, at number 6 of 10, indicating room for improvement. Table 18 shows the ranking of mobile apps from highest to lowest for Player Information.

Table 18. Expert Ratings: Player Information

Expert Ratings: Player Information	
Application	Average Cluster Rating
Real Madrid	3.79
Onefootball	3.64
FC Barcelona	3.58
Seahawks	3.45
Brooklyn Nets	3.36
Sounders FC	**3.35**
Arsenal	3.13
FC Bayern Munich	3.04
Chicago Fire	2.79
ESPN FC	2.17
Average rating	3.23
Std dev	0.48

For Schedule, Results and League, the average expert Likert rating for all applications in the study was 3.45. The Sounders FC application was 3.25, as shown in Table 19. Again, the Sounders FC application was in the bottom half of the rated applications for this scenario, at number 6 of 10.

Table 19. Expert Ratings for Schedule, Results and League

Expert: Schedule, Results & League	
Application	Average Cluster Rating
FC Barcelona	4.10
Onefootball	3.95
Seahawks	3.83
Brooklyn Nets	3.72
Real Madrid	3.38
Sounders FC	**3.25**
Arsenal	3.25
ESPN FC	3.19
FC Bayern Munich	2.98
Chicago Fire	2.89
Average rating	3.45
Std dev	0.42

TEDS User Ratings

Although the expert ratings were collected for all applications in all scenarios, only the Player Information and Schedule, Results and League scenarios were collected from the purposive sampled users from Facebook for the Sounders FC app. Thus, this paper presents only the expert ratings in the tables, except for when the table states that the ratings include the user ratings, such as the findings related to RQ4. Those ratings, which are discussed at the end of this chapter, compared the Sounders FC app ratings between expert and purposive sampled users for those two scenarios. (See Appendix K for the full list of Phase 1 user ratings of Player Information and Schedule, Results, and League for the Sounders FC iOS app.)

Phase 2: Treatment Development

Phase 2 of this project addresses RQ2 to develop a method to generate feature improvement specifications and to develop feature treatments for the Sounders FC mobile application. Details of the RQ2 findings follow this current section about RQ1. However, for the purpose of discussing RQ1, four treatments were generated for the Phase 3 A/B testing phase:

- Add a large photo and horizontal scrolling to the roster page
- Show player statistics in a graph.
- Add a tab that opens career statistics on a new page
- Link to the online store to purchase a jersey from the player card

Phase 3: A/B Testing and Iteration

The Sounders FC development organization was presented with the Player Information feature improvement specification from Phase 2, which included the four recommended treatments. However, one of the developers scheduled to work on features for this study had left the Sounders development team, and there was limited time to complete the A/B experiments since the season was ending and the A/B tests needed to be completed before the usage patterns of the app changed in the off-season. Although the Sounders hired a new developer to replace the attrition, some ramp-up time would be required for the new developer. Therefore, the Sounders FC chose to A/B test two of the four recommended player card treatments:

- Add a large photo and horizontal scrolling to the roster page
- Show player statistics in a graph

The experiments ran simultaneously and shared the same OEC: the average number of player cards selected by each user in each experiment. The event tracked in Flurry was called *Player Detail – Visible*. The assumption was that users who find the player information they want click the player cards more often over time if the application met their needs better. The event fired each time a player card was selected.

This A/B experiment was tracked for three weeks (October 3-24, 2016). Table 24 shows the experiment results. The Active Users column includes users who had navigated to the roster page.

The Events column includes the number of times a user then navigated to a player card. Users with the treatment saw a large photo of a player, with a horizontal listing of roster players at the bottom of the page. Users with the control saw a vertical listing of roster players, with no large player photo. The treatment for this experiment is described in detail later in this chapter in the RQ2 discussion.

Table 24. Results for Add a Large Photo and Horizontal Scrolling to the Roster Page

Results for Add a Large Photo and Horizontal Scrolling to the Roster Page						
	OEC = "Player Detail – Visible" Event					
	Control			Treatment		
Date	Active Users	Unique Users with Event	Events	Active Users	Unique Users with Event	Events
10/3/2016	147	19	103	126	14	24
10/4/2016	104	3	3	101	11	29
10/5/2016	133	9	45	113	9	48
10/6/2016	145	7	11	123	10	20
10/7/2016	171	10	35	152	11	68
10/8/2016	159	12	60	144	7	12
10/9/2016	156	11	46	142	11	60
10/10/2016	144	5	91	154	12	46
10/11/2016	197	11	115	167	10	30
10/12/2016	**541**	**115**	**449**	**573**	**232**	**737**
10/13/2016	280	17	112	259	11	26
10/14/2016	170	5	17	194	9	16
10/15/2016	239	14	62	233	17	48
10/16/2016	**556**	**31**	**117**	**606**	**90**	**264**
10/17/2016	209	7	14	193	6	27
10/18/2016	183	10	75	187	14	46
10/19/2016	176	7	17	180	8	27
10/20/2016	169	4	10	159	7	20
10/21/2016	196	11	16	183	6	12
10/22/2016	247	10	49	247	19	65
10/23/2016	**757**	**85**	**271**	**765**	**166**	**449**
10/24/2016	303	9	58	302	18	98
Total	**5382**	**412**	**1776**	**5303**	**698**	**2172**
Avg per day	**244.64**	**18.73**	**80.73**	**241.05**	**31.73**	**98.73**
Avg total events per user			**0.3300**			**0.4096**
Pct UU with event		**0.0766**			**0.1316**	

Experiment		**A Conv =**	**0.3300**		**std error A**	**0.0064094**
		B Conv =	**0.4096**		**std error B**	**0.0067529**
zscore =	**8.5486**		**Confidence**	**= 99.99%**		
*UU - Unique Users						

During this A/B experiment, the average number of player card viewings with the control was 0.3300. The average number of viewings of the player card with the treatment was 0.4096. This means that on average, users with the treatment were more likely to interact with the roster page by clicking the player card than users with the control, as shown by measuring the Player Detail – Visible event in this experiment.

A *zscore* (or standard score) is a standardized value based on a random variables value, the population mean, and the standard deviation (Pullen, 2010). It is used in A/B testing to determine statistical significance in an experiment where there is an assumption of a normal distribution of the random variable, which is the case in this experiment. Larger zscores indicate that the variable is further away from the mean of the distribution. A zscore greater than 2 means that the variable is 2 standard deviations from the mean, which corresponds to 95% confidence that the difference was not due to random chance. In this experiment, the zscore of 8.5486 indicated a 99.99 percent confidence that the difference between the control and the treatment was due to the differences in the new treatment. In other words, the data showed with high confidence that, on average, users with the treatment clicked player card more often than the users with the control did.

The difference between the treatment and the control is pronounced on game days during the experiment (10/12, 10/16, and 10/23). This makes sense, since users are more likely to use the Sounders FC mobile app during a game. At the end of this experiment, the recommendation was to deploy the treatment version of the player card because it had won the A/B experiment.

Table 25 shows the results of the next A/B experiment: Show player statistics in a graph. This treatment is also described in detail later in this chapter when RQ2 is discussed. The treatment for this experiment was on the player card. Users with the treatment saw the player statistics in graphical form on the player card. Users with the control saw the player statistics listed in static text in a table format on the player card.

The average number of player card viewings for users with the control was 0.0516. The average number of player card viewings with the treatment was 0.0529. With a zscore of .8866, this maps to only 65.59 percent confidence that the difference between the control and the treatment was due to the differences in the new treatment. Even though the OEC is slightly improved in the data, the zscore does not show significance above 95% confidence, which means the treatment did not win the experiment.

Table 25. Results for Show Player Statistics in a Graph

Results for Show Player Statistics in a Graph						
	OEC = "Player Detail – Visible" Event					
	Control			Treatment		
Date	Active Users	Unique Users with Event	Events	Active Users	Unique Users with Event	Events
10/3/2016	1951	31	139	2045	20	40
10/4/2016	1426	8	8	1450	17	35
10/5/2016	1538	15	60	1703	15	57
10/6/2016	1388	10	39	1441	15	39
10/7/2016	1519	12	38	1573	18	77
10/8/2016	1544	12	60	1568	8	13
10/9/2016	1475	12	48	1537	17	77
10/10/2016	1421	8	118	1545	15	60
10/11/2016	1848	14	121	1946	17	42
10/12/2016	**5636**	**182**	**628**	**5653**	**287**	**865**
10/13/2016	1948	19	115	1980	16	34
10/14/2016	1319	6	18	1405	14	26
10/15/2016	1847	20	82	1883	18	59
10/16/2016	**4589**	**58**	**204**	**4599**	**103**	**291**
10/17/2016	1524	10	19	1602	8	29
10/18/2016	1357	14	86	1439	17	51
10/19/2016	1278	10	40	1288	10	36
10/20/2016	1256	7	16	1296	8	21
10/21/2016	1490	13	22	1467	9	16
10/22/2016	2031	11	51	2109	23	75
10/23/2016	**6674**	**136**	**427**	**6705**	**196**	**515**
10/24/2016	2370	21	109	2455	27	117
Total	**47429**	**629**	**2448**	**48689**	**878**	**2575**
Avg per day	**2155.86**	**28.59**	**111.27**	**2213.14**	**39.91**	**117.05**
Avg total events per user			**0.0516**			**0.0529**
Pct UU with event		**0.0133**			**0.0180**	
Experiment		**A Conv =**	**0.0516**		**std error A**	**0.0010159**
		B Conv =	**0.0529**		**std error B**	**0.0010143**
zscore =	**0.8866**		**Confidence**	**= 65.59%**		
*UU - Unique Users						

In this experiment, Active Users in Table 25 refers to all individuals who used the Sounders FC mobile app on a specific day. For example, on October 12, 5,636 users saw the control, and 5,653 saw the treatment. Events in the table refers to each time the control or the treatment player card was shown. So, on October 12,628 users saw the control version event, and 865 saw the treatment version. This experiment included all application users, even though the treatment version of the roster page was seen by relatively very few users. It would have been preferable to include in the experiment only users who interacted with the roster page. This would have resulted in higher confidence numbers (all other things being equal). However, since the assignment of users between these two simultaneously running experiments was not random, it was not possible to reduce the number of users in this experiment without biasing the result of the experiment with the horizontal scroll on the roster page.

The Sounders FC organization chose to deploy the graphical representation of the player statistics treatment to all users, even though the experiment did not suggest with scientific confidence that it was a measurable improvement. One reason was the Sounders FC iOS app already showed game statistics graphically in the game summary feature, so they wanted the player statistics to be shown graphically for consistency.

Phase 4: Result Comparison

It is important to compare the Phase 1 and the Phase 4 user and expert ratings to see how the ratings changed after the recommended treatments were deployed. Table 30 presents this comparison for Player Information.

Table 30. Comparison of Phase 1 and Phase 4 Ratings for Player Information

Comparison of Phase 1 and 4 Ratings for Player Information									
	Expert			User			Averaged		
Cluster	Phase 1 Expert	Phase 4 Expert	Difference	Phase 1 User	Phase 4 User	Difference	Phase 1 Average	Phase 4 Average	Difference
Navigation and Findability	4.00	4.10	0.10	3.46	3.71	0.24	3.73	3.90	0.17
Structure	3.60	4.15	0.55	3.59	3.84	0.25	3.59	4.00	0.40
Identity	3.50	3.90	0.40	3.96	3.97	0.01	3.73	3.94	0.21
Parsimony	3.60	3.70	0.10	3.68	3.67	-0.01	3.64	3.69	0.05
Completeness	3.20	3.60	0.40	3.91	3.87	-0.04	3.55	3.73	0.18
Trustworthiness	3.70	3.80	0.10	4.05	4.15	0.09	3.88	3.97	0.10
Interaction	3.30	3.50	0.20	3.34	3.26	-0.08	3.32	3.38	0.06
Customization	2.60	2.70	0.10	3.50	3.48	-0.02	3.05	3.09	0.04
Savings	3.00	3.10	0.10	3.63	3.68	0.05	3.32	3.39	0.08
Confidence	3.30	3.40	0.10	3.80	4.11	0.31	3.55	3.76	0.20
Attractiveness	3.20	3.55	0.35	3.61	4.10	0.49	3.40	3.82	0.42
Enjoyment	3.20	3.60	0.40	3.45	3.71	0.26	3.33	3.65	0.33
Total averages	3.35	3.59	0.24	3.67	3.80	0.13	3.51	3.69	0.19

The first three columns compare the expert ratings in Phase 1 and Phase 4; the third column shows the difference. The next three columns compare the user ratings, and the final three columns show the average of the expert and user ratings for each phase. The average change between Phase 1 and Phase 4 for the expert ratings was 0.24. The average change for user ratings was 0.13. The user raters were included in this table for comparison purposes. The user ratings were all from unique raters in both phases. There could be concern about bias with the expert raters, since the same expert raters participated in Phase 1 and Phase 4. However, only one expert rater had detailed knowledge of the changes that had occurred between phases.

Table 31 shows the difference between Phase 1 and Phase 4 in average TEDS ratings for the Schedule, Results, and League scenario. This table is included to show the detailed improvement in the expert and user ratings for this other scenario that had user ratings. The average change between the two phases for the expert ratings was 0.35. The average change for user ratings was 0.33. This increase in average rating is larger than the 0.13 increase seen in the Player Information scenario. Table 32 shows the average rating changes between Phase 1 and Phase 4 for all scenarios.

Table 31. Comparison of user and expert ratings in Phase 1 and Phase 4 for Schedule, Results, and League

Comparison of Phase 1 and 4 Ratings for Schedule, Results and League									
	Expert			User			Averaged		
Cluster	Phase 1 Expert	Phase 4 Expert	Difference	Phase 1 User	Phase 4 User	Difference	Phase 1 Average	Phase 4 Average	Difference
Navigation and Findability	3.90	4.00	0.10	3.68	3.93	0.25	3.79	3.96	0.18
Structure	3.60	3.80	0.20	3.45	3.81	0.36	3.53	3.80	0.28
Identity	3.60	4.00	0.40	3.68	3.93	0.25	3.64	3.96	0.33
Parsimony	3.30	3.50	0.20	3.68	3.86	0.18	3.49	3.68	0.19
Completeness	3.10	4.10	1.00	3.71	3.95	0.24	3.40	4.03	0.62
Trustworthiness	3.60	4.00	0.40	3.90	4.17	0.27	3.75	4.09	0.33
Interaction	2.30	3.10	0.80	3.17	3.33	0.16	2.73	3.21	0.48
Customization	2.40	2.50	0.10	3.39	3.61	0.22	2.89	3.05	0.16
Savings	3.40	3.50	0.10	3.48	3.95	0.47	3.44	3.73	0.28
Confidence	3.00	3.10	0.10	3.81	4.27	0.46	3.40	3.68	0.28
Attractiveness	3.40	3.75	0.35	3.48	4.10	0.62	3.44	3.93	0.48
Enjoyment	3.40	3.85	0.45	3.42	3.85	0.43	3.41	3.85	0.44
Total averages	3.25	3.60	0.35	3.57	3.90	0.33	3.41	3.75	0.34

Table 32 shows the change in expert ratings for each of the eight scenarios. The ratings improved for almost every cluster for all scenarios from the first assessment to the second assessment except for In-Seat Concessions. In-Seat Concessions was not included in this comparison because the other iOS mobile sports apps in this research did not support it. (See Appendix N for a detailed comparison of Phase 1 and Phase 4 evaluations for the other six scenarios.)

Table 32. Change in Expert Ratings from Phase 1 to Phase 4 for Each Cluster

Change in Expert Ratings from Phase 1 to Phase 4 for Each Cluster										
Clusters	**Player Information**	**Schedule, Results and League**	**TeamNews**	**Merchandise and Store**	**Ticketing**	**Social Media Integration**	**Game Day and Stadium Guide**	**In-Seat Concessions**	**Std Dev**	**Average**
Navigation and Findability	0.10	0.10	0.30	0.00	2.10	0.30	1.80	0.00	0.85	0.59
Structure	0.55	0.20	0.00	0.30	2.20	0.50	1.60	0.00	0.80	0.67
Identity	0.40	0.40	0.10	0.10	2.20	0.30	1.50	0.00	0.79	0.63
Parsimony	0.10	0.20	0.10	0.10	1.90	0.30	1.50	0.00	0.74	0.53
Completeness	0.40	1.00	0.20	0.10	2.30	0.20	1.40	0.00	0.81	0.70
Trustworthiness	0.10	0.40	0.10	0.10	2.40	0.20	1.90	0.00	0.94	0.65
Interaction	0.20	0.80	0.40	0.10	1.90	0.20	1.20	0.00	0.66	0.60
Customization	0.10	0.10	0.00	0.30	1.70	0.00	1.30	0.00	0.67	0.44
Savings	0.10	0.10	0.20	0.10	2.00	0.30	1.50	0.00	0.77	0.54
Confidence	0.10	0.10	0.00	0.00	1.60	0.00	1.90	0.00	0.80	0.46
Attractiveness	0.35	0.35	0.20	0.20	1.90	0.20	1.70	0.00	0.74	0.61
Enjoyment	0.40	0.45	0.20	0.20	1.60	0.30	1.60	0.00	0.64	0.59
Average Change	0.24	0.35	0.15	0.13	1.98	0.23	1.58	0.00	0.75	0.67

Change in Scenario TEDS Ratings Between Phase 1 and Phase 4

It is worth comparing the change in expert TEDS scenario ratings for the Sounders FC iOS mobile application between Phase 1 and Phase 4. After completing A/B testing of two Player Information treatments, the Sounders FC development team deployed both recommended treatments from Phase 2 to all users in Phase 3. Thus, the Phase 4 expert TEDS ratings assessed the updated feature set for this scenario. The average expert cluster rating was 0.24 higher in Phase 4 than in Phase 1, which was significantly lower than the average rating improvement for all scenarios (0.67). Table 33 shows the average change in expert cluster ratings between Phase 1 and Phase 4 for each scenario.

Table 33. Change in expert ratings between Phase 1 and Phase 4 for each scenario

Scenario rating changes										
	Player Information	**Schedule, Results and League**	**TeamNews**	**Merchandise and Store**	**Ticketing**	**Social Media Integration**	**Game Day and Stadium Guide**	**In-Seat Concessions**	**Std deviation**	**Average**
Average Change	0.24	0.35	0.15	0.13	1.98	0.23	1.58	0.00	0.75	0.67

On the surface, the data appears to suggest that the iterative process from this dissertation was not effective because the one scenario in which the process was applied did not improve the TEDS rating above the average for all scenarios. However, additional relevant information explains the increases in ratings for the other scenarios. Table 34 shows a comparison of the average cluster ratings in Phase 1 and Phase 4 for the eight scenarios.

Table 34. Comparison of Average Cluster Ratings in Phase 1 and Phase 4 for Each Scenario

	Sounders FC Expert Scenario Ratings									
Values	PlayerInfo	Schedule, Results and League	TeamNews	Merchandise and Store	Ticketing	Social Media Integration	Game Day and Stadium Guide	In-Seat Concessions	Std Dev	Average
Phase 1 Average Cluster Rating	3.35	3.25	3.36	3.20	**1.20**	2.74	**1.00**	**1.00**	1.11	2.39
Phase 4 Average Cluster Rating	3.59	3.60	3.51	3.33	3.18	2.98	2.58	**1.00**	0.87	2.97

In the first phase of assessments, recall that the Sounders iOS mobile application was in Beta form. The Ticketing scenario and Game Day and Stadium Guide scenario were not fully supported when the Phase 1 expert evaluations occurred. When the Phase 4 evaluation was performed, the Sounders application had deployed numerous features so it was no longer a Beta version. The average improvement in Ticketing (1.98) and in Game Day and Stadium Guide (1.58) are due to these scenarios being removed from the Beta version and enabled in the application. However, these two scenarios still lagged behind Player Information, Schedule, Results and League, Team News, and Merchandise and Store scenarios. The improvement for these scenarios was larger than the other scenarios, but the magnitude is likely due to the lack of (or little) availability in the application when it was assessed in Phase 1.

The next question is what explains the increase in ratings for the other scenarios? And why did the

Schedule, Results, and League scenario improve by 0.35 when the Sounders FC development team did not have time to implement the recommended treatments from Phase 2? The development team continued to deploy features throughout the period between the Phase 1 and Phase 4 TEDS assessments. The two recommended Player Information treatments were included in these deployments, but the Sounders had an existing backlog of items that were prioritized above the Player Information treatments implemented in Phase 2.

Tracking Development Changes Between Phase 1 and Phase 4

This research study was embedded in a live development project, thus many development feature deployments occurred to the Sounders FC iOS mobile application between Phase 1 and Phase 4. Fortunately, the development team tracked feature improvement items that were included in each deployment between Phase 1 and Phase 4. Table 35 presents the first five development items in the Sounders iOS mobile application backlog between Phase 1 and Phase 4. (For the full list of 57 work items, see Appendix O.)

Each row in Table 35 corresponds to a feature improvement item. The Item column contains the index of deployed feature items. The Version column corresponds to the deployment version number. The Work Item column corresponds to the feature improvement item name. The Scenario and Cluster columns were added to track how each feature improvement item related to this TEDS study. Each work item was analyzed tagged with information about the TEDS scenarios and clusters that would likely be affected by the development item.

Table 35. First Five Development Items in the Sounders FC iOS Application Backlog

Item	Version	Work Item	Scenario	Cluster
1	3.5.5	Updated Match Details screen	Schedule, Results and League	Structure, Completeness, Attractiveness, Enjoyment
2	3.5.5	Updated MatchDay Guide screen	Schedule, Results and League	Structure, Attractiveness
3	3.5.5	Various UI adjustments and bug fixes	TBD	TBD
4	3.5.3 (1809)	Standings layout updated to include the Supporters' Shield	Schedule, Results and League	Navigation and Findability, Structure, Completeness, Savings, Enjoyment
5	3.5.3 (1809)	Fixed bug related to news push notifications	TeamNews	Completeness, Savings

In the table, the first, second, and fourth items affected the Schedule, Results, and League scenario. It is unclear which scenarios were affected by the third item. The fifth item affected the TeamNews scenario. The Cluster column identifies the TEDS clusters that the development item was expected to affect.

With the work items tagged with scenarios and clusters affected, it was possible to add the number of line items for each scenario and cluster to quantify the changes that occurred in the Sounders FC iOS application between Phase 1 and Phase 4. This method was imprecise, since there is no weighting given to the impact of individual line items, and certainly some line items are much more impactful. Also, the analysis of mapping line items to the scenarios and clusters was completed by the researcher with only auditing by the Sounders FC development team. In addition, some feature-enabling work items that were initially in Beta form (like Ticketing) do not appear in the list at all. Nevertheless, this method shows the rough development activity by the Sounders FC development team by scenario and cluster between Phase 1 and Phase 4.

Table 36 shows the estimated count of work items (bottom row) as mapped by the researcher. The Sounders FC development team did not provide estimated costs for the work to make the non-beta feature visible to users. For example, it is likely that the Ticketing feature was almost completed in Phase 1, but it got rated as 1.20 because it was not visible. This feature had a few very minor features available, but users could not see or purchase tickets. Similarly, Game Day and Stadium Guide features (which were rated 1.0) were not available in Phase 1. In Phase 4, users were able to interact with these features because the Sounders had enabled both features and removed the Beta status. This undoubtedly accounts for the large increases in the ratings between Phase 1 and Phase 4 for both scenarios.

Table 36. Change between Phase 1 and Phase 4 ratings for each cluster by scenario

Sounders FC expert scenario rating changes										
Values	**Player Information**	**Schedule, Results and League**	**TeamNews**	**Merchandise and Store**	**Ticketing**	**Social Media Integration**	**Game Day and Stadium Guide**	**In-Seat Concessions**	**Std deviation**	**Average**
Navigation and Findability	0.10 8, 27	0.10 4,8,16,17	0.30 6,8,16,17	0.00	2.10 23, 54	0.30 53	1.80 27,51	0.00	0.85	0.67
Structure	0.55 8	0.20 1,2,4,8,17	0.00 6, 8, 11	0.30	2.20 31	0.50	1.60 51	0.00	0.80	0.76
Identity	0.40 27, 50	0.40 17	0.10	0.10	2.20 23, 43, 49	0.30	1.50 27	0.00	0.79	0.71
Parsimony	0.10 8	0.20 8,17,53	0.10	0.10	1.90	0.30 53	1.50	0.00	0.74	0.60
Completeness	0.40 27, 55	1.00 1,4,12,13,1 4,29,47	0.20 5, 36,27,42,46	0.10	2.30 23, 38, 40, 56	0.20 53	1.40 27,35,38,42 ,47,51,57	0.00	0.81	0.80
Trustworthiness	0.10 33	0.40 12,13,14,47	0.10 19, 42, 46	0.10	2.40 54	0.20	1.90 42, 51	0.00	0.94	0.74
Interaction	0.20	0.80 16	0.40	0.10	1.90	0.20	1.20 39,47	0.00	0.66	0.69
Customization	0.10	0.10	0.00	0.30	1.70 43,44	0.00	1.30	0.00	0.67	0.50
Savings	0.10 8, 55	0.10 4, 8, 29	0.20 5,8,11,37,4 2,46	0.10	2.00 38,54,56	0.30	1.50 30,35,38,42 ,51, 57	0.00	0.77	0.61
Confidence	0.10 33	0.10	0.00 36,42,46	0.00	1.60	0.00	1.90 42	0.00	0.80	0.53
Attractiveness	0.35 32, 33, 50	0.35 1,2,16,27,2 9,53	0.20 6, 18, 42	0.20	1.90 31	0.20	1.70 39,42,51	0.00	0.74	0.70
Enjoyment	0.40 27, 32, 33, 50	0.45 1,16,21,29, 53	0.20 18,19,36, 37, 46	0.20	1.60	0.30 53	1.60 27,35,51, 57	0.00	0.64	0.68
Average Change	0.24	0.35	0.15	0.13	1.98	0.23	1.58	0.00	0.77	0.67
Count of work items	**19**	**40**	**32**	**0**	**15**	**4**	**28**	**0**	**Total**	**138**

The smaller numbers in the table refer to the line items in the Sounders FC development team's work items list, which are partially shown in table 35. The count of work items indicate the amount of work that occurred in each scenario. For the remaining six scenarios, a correlation appeared to exist between the number of work items and the increase in scenario ratings. The count of work items (40) for Schedule, Results, and League, coupled with its improvement by 0.35, suggests that this scenario received the most development attention between assessments. TeamNews improved by 0.15, with 32 development work items correlated. Although Social Media Integration improved by 0.23, with only four tagged development work items, it was also in Beta form during Phase 1. Its improvement can be attributed to some minor features being enabled by Phase 4.

The following paragraph is a summary of the RQ1 findings.

RQ1 Findings Summary

This study demonstrated that an organization can use the TEDS assessment method to evaluate an information artifact in relation to similar artifacts to generate a context-specific rating for comparison. The TEDS ranked assessment data and corresponding evaluation comments and screenshots can then be used to generate treatment ideas to address weaknesses in the information artifact for use in A/B testing. A/B testing is a useful evaluation method for providing a quantitative way to triangulate the analysis criteria of survey data.

Addressing Research Question 2 (RQ2)

How can the output from a TEDS analysis be used to generate feature ideas that map to improvements in the TEDS assessment of an information artifact?

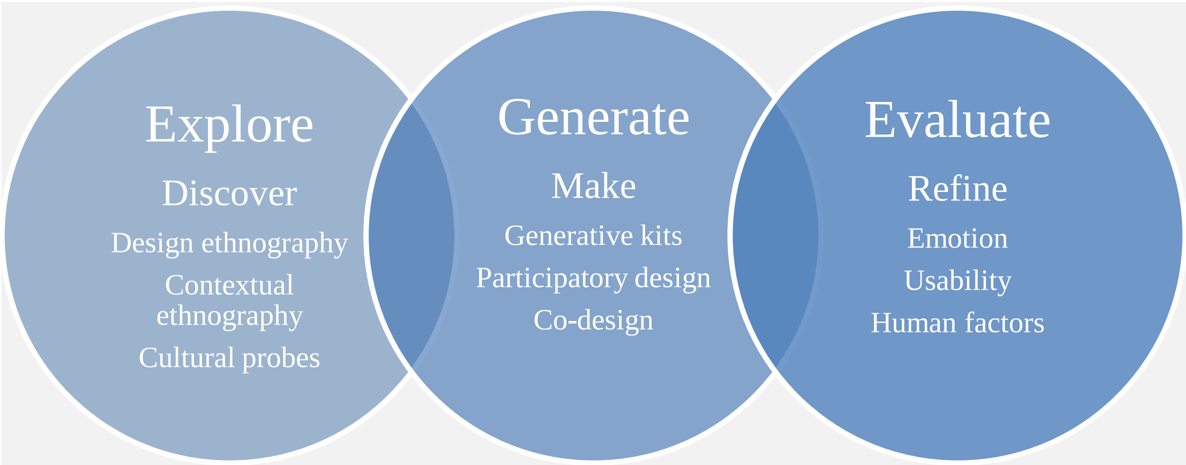
Phase 2: Treatment Development Phase

The first goal of the treatment development phase was to develop a systematic method to generate feature improvement specifications based on TEDS assessments generated in Phase 1. The Delphi method was used for this purpose. The second goal was to use the developed method to produce treatment improvement specifications for the Seattle Sounders FC organization for the treatment development and the A/B testing phases.

Delphi Focus Group Findings

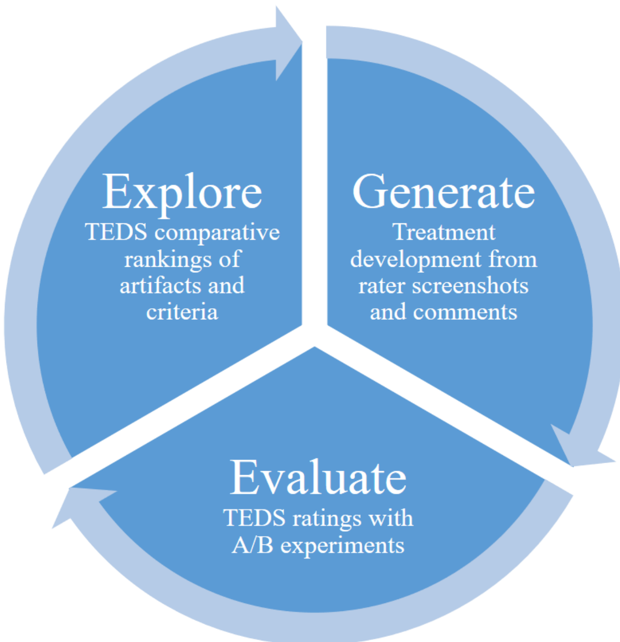
The first question in the focus group questionnaire asked each participant to write about the process they used or would use when tasked with generating design ideas. The responses described design phases, tools they used in the design phases, and techniques to generate ideas. During the corresponding focus group discussion, participants chose to leverage the Exploratory, Generative, and Evaluative phases from the Carnegie Mellon design education method (Hanington, 2010). Because these phases are characterized by approaches rather than specific methods, the diverse ideas from participants all fit in at least one of the phases. This approach became a useful guide for categorizing and describing the design steps that occurred in this study. Figure 8 shows the Carnegie Mellon Design model (Hanington, 2010).

Figure 8. Model of Design Research (Carnegie Mellon Design Education Method)



This study employs numerous design approaches that align with the Carnegie Mellon design phases. The Carnegie Mellon model is also iterative, in that the evaluation phase informs the exploration phase, which informs the generation phase to promote continuous improvement. Figure 9 shows how the different approaches used in this dissertation align with the Carnegie Mellon design phases.

Figure 9. Design approaches used and their alignment with Carnegie Mellon design phases



Evaluate Design Phase

The Carnegie Mellon model includes both explorative techniques to inform generative approaches and evaluative tools to determine the effectiveness of designs. A design can start at any phase of the Carnegie Mellon model. This study begins in Phase 1 at the evaluation stage by using the TEDS method to assess the 10 iOS mobile applications. The ratings produced from the TEDS method measure specific characteristics of each artifact, which enables ranking and comparison. The assessment results provide insight into which TEDS attributes are meaningful for each scenario.

Phase 3 adds a quantitative method of evaluation with A/B testing. Unlike the TEDS assessment method, which provides qualitative evaluations to compare performance of more than one IT artifact, A/B testing evaluates treatment modifications made to only one artifact at a time by comparing web statistics to see the impact from changes made to the artifact.

Explore Design Phase

Exploratory elements are included in Phase 2. The TEDS ratings paired with user comments and screenshots for each artifact shed light on designs that better meet the user's needs for a given attribute in a specific context. This pairing also provides a way to identify best practices for the artifacts in the study. The TEDS ratings can be pivoted by average scenario rating or by average cluster rating, which makes it possible to rank the corresponding comments and screenshots.

To inform their scenario selection, the Sounders FC development team used the average scenario ratings by the expert raters and their own understanding of the relative significance of each scenario. A cluster ranking shows how an artifact compares to other artifacts for the criteria for all scenarios. Like scenario ratings, cluster ratings can be used to compare artifacts, but it can also show negligible differences between all artifacts for a given cluster criteria (indicated by a small standard deviation). Large differences in the average cluster rating for an artifact would suggest a systematic design issue for the artifact.

Table 22 shows TEDS average expert cluster ratings for Player Information for all 10 iOS mobile applications. With a score of 3.35, the Sounders FC application scored slightly above the average of the 10 applications (3.23). The Real Madrid and Onefootball applications achieved the highest average ratings, with 3.79 and 3.64 respectively (indicated in bold black in the bottom row). The design patterns in higher-rated artifacts (whether scenario or cluster ratings) can be used to generate treatment ideas for the treatment improvement specifications and to prioritize data to explore for feature ideas and design choices.

Table 22. Phase 1: TEDS Expert Cluster Ratings for Player Information

Phase 1: TEDS Expert Cluster Ratings for Player Information												
	iOS Application											
Clusters	Sounders FC	Seahawks	FC Barcelona	Real Madrid	Chicago Fire	Bayern	Onefootball	ESPN FC	Brooklyn Nets	Arsenal FC	Average rating	Std dev
Navigation and Findability	4.00	3.65	3.70	3.70	3.20	3.50	3.90	**1.90**	3.80	3.50	3.44	0.61
Structure	3.60	3.60	**4.00**	3.80	3.10	3.00	3.90	**2.30**	3.90	3.50	3.41	0.56
Identity	**3.50**	**3.64**	**3.60**	**4.00**	3.40	**3.00**	**3.80**	**2.40**	**3.70**	3.40	3.42	0.48
Parsimony	3.60	3.33	**3.80**	3.50	2.90	**2.60**	3.60	**2.10**	2.80	3.00	3.18	0.54
Completeness	3.20	3.87	3.40	**4.00**	**2.90**	3.40	**3.90**	**2.40**	3.68	3.40	3.38	0.52
Trustworthiness	3.70	**4.67**	4.20	3.90	3.40	3.70	3.80	**2.30**	3.60	3.70	3.71	0.64
Interaction	3.30	2.70	**3.70**	**3.80**	**1.90**	2.90	**3.60**	**1.80**	2.50	2.70	2.96	0.74
Customization	2.60	2.93	2.80	**3.10**	2.30	2.40	**3.30**	**2.00**	2.30	2.20	2.68	0.45
Savings	3.00	3.20	3.20	**3.80**	2.60	3.00	3.20	**2.00**	**3.60**	2.70	3.00	0.54
Confidence	3.30	3.30	3.30	3.40	**2.60**	3.40	3.40	**2.20**	3.30	3.20	3.11	0.43
Attractiveness	3.20	3.30	3.80	**4.40**	**2.60**	2.80	3.70	**2.40**	3.70	3.30	3.28	0.64
Enjoyment	3.20	3.27	3.40	**4.10**	**2.60**	2.80	3.60	**2.20**	3.40	3.00	3.15	0.57
Average Likert	3.35	3.45	3.58	**3.79**	2.79	3.04	**3.64**	**2.17**	3.36	3.13	3.23	3.22

The fact that an information artifact does poorly in a scenario does not mean that making modifications to improve its performance is the highest priority. For example, most of the mobile applications do not support the In-Seat Concessions scenario. (In this scenario, the application enables game attendees to order food from their seat and either have it delivered or ready for pickup at the concession stand.) Although In-Seat Concessions is novel, implementing it is likely less important than making competitive improvements to the Player Information, Team News, or Schedule, Results, and League feature sets. The Sounders FC confirmed this, stating that the Player Information scenario and the Schedule, Results and League scenario had higher business impact for them. Table 23 shows the average expert cluster ratings for all scenarios for the 10 iOS applications.

Table 23. Phase 1: Average Expert Cluster Ratings for All Scenarios

Phase 1: Average Expert Cluster Rating for All Scenarios												
	iOS Application											
Clusters	**Sounders FC**	**Seahawks**	**FC Barcelona**	**Real Madrid**	**Chicago Fire**	**Bayern**	**Onefootball**	**ESPN FC**	**Brooklyn Nets**	**Arsenal FC**	**Average rating**	**Std dev**
Navigation and Findability	3.44	3.35	**3.77**	3.30	**2.85**	3.05	**3.63**	**2.73**	3.56	3.15	3.28	0.34
Structure	3.34	3.43	**4.03**	3.50	**2.83**	3.00	**3.83**	**2.75**	3.62	3.18	3.35	0.42
Identity	3.42	3.45	3.88	3.48	**2.91**	**2.87**	3.75	**3.00**	3.58	3.07	3.34	0.36
Parsimony	3.18	3.35	**3.68**	3.22	**2.80**	2.90	**3.58**	**2.75**	3.32	2.97	3.17	0.32
Completeness	3.26	3.75	**3.91**	3.80	**2.85**	**3.03**	**3.85**	**2.98**	3.59	3.22	3.42	0.40
Trustworthiness	3.50	**3.95**	**4.13**	3.70	**3.13**	3.38	3.88	**3.00**	3.65	3.48	3.58	0.36
Interaction	2.86	2.92	**3.55**	3.28	**2.58**	**2.48**	**3.58**	2.68	2.99	2.68	2.96	0.39
Customization	2.58	2.98	**3.27**	2.88	**2.44**	2.52	**3.40**	2.55	2.80	2.55	2.80	0.33
Savings	3.10	3.23	**3.62**	3.33	**2.76**	**2.82**	**3.68**	**2.70**	3.33	2.93	3.15	0.35
Confidence	3.10	3.16	**3.45**	3.32	**2.74**	3.13	**3.48**	**2.80**	3.28	3.13	3.16	0.24
Attractiveness	3.24	3.31	**4.04**	3.63	**2.60**	2.87	**3.73**	**2.73**	3.45	3.12	3.27	0.46
Enjoyment	3.14	3.14	**3.87**	3.50	**2.58**	**2.75**	**3.85**	**2.70**	3.33	3.00	3.18	0.45
Average Likert	3.18	3.34	**3.77**	3.41	**2.75**	2.90	**3.68**	**2.78**	3.37	3.04	3.22	0.36

For each scenario, the cluster criteria should be analyzed to determine whether each cluster is germane to the scenario. For example, the Confidence cluster measures whether the artifact provides enough security and safety to make users feel comfortable. The table shows that the standard deviation for Confidence is 0.24, which is the smallest difference of all clusters in this study, indicating negligible differences among all artifacts for this cluster. Although Confidence is worth including in the TEDS framework, when analyzing Player Information, Team News, or Schedule, Results and League, little difference exists between the artifacts. Furthermore, little security or safety risk exists for users consuming sports information on the mobile applications. Removing less significant clusters like Confidence for a scenario helps prioritize the clusters that make a substantial difference and reduces the amount of data to consider.

Generate Design Phase

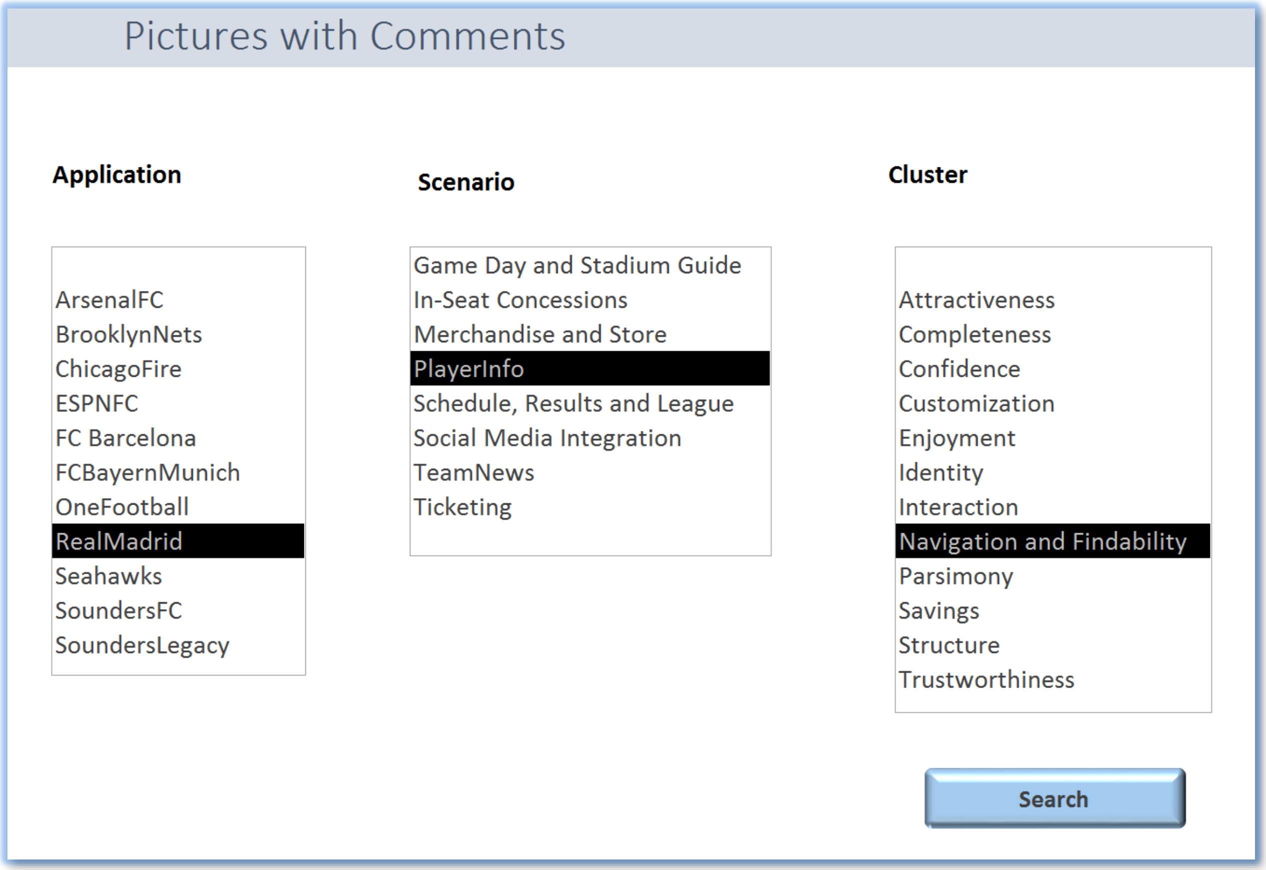
Generative elements are included in Phase 2, when the TEDS ratings, comments, and screenshots are analyzed to generate treatment improvement ideas. Focus group participants recommended competitive analysis and user feedback as procedures to help generate ideas. The rater comments and screenshots provide concrete examples of designs from other artifacts that meet a user's needs for a specific scenario. For example, in this study there needed to be treatment ideas generated for the Player Information scenario. The comments and screenshots from the raters of the iOS mobile apps that ranked highly for this scenario (like the Real Madrid or Onefootball apps), provided examples of design patterns that helped brainstorm treatment ideas for the Sounders iOS mobile app.

Despite gaps in the comments and screenshots for some artifacts, the task of searching through the rater comments and screenshots was daunting. The TEDS comments and screenshots were compiled in a Microsoft Access database, which included more than 25,000 rows of expert rating assessment data. To analyze the comments and screenshots expediently, a search tool was built to filter the comments and screenshots by artifact, scenario, and cluster. In an ideal TEDS data set, at least several rater comments and screenshots would exist for each cluster and scenario.

TEDS Rating, Comments, and Screenshot Search Tool

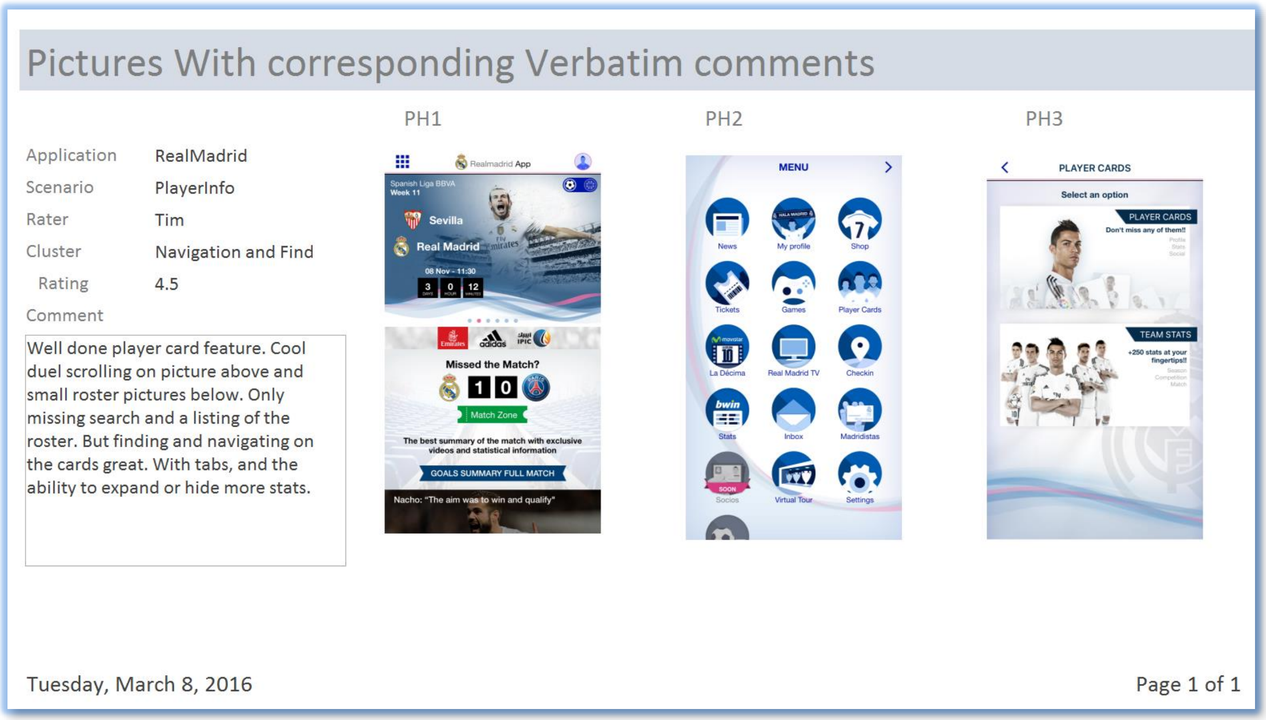
Figures 10 and 11 show screenshots of the search tool that was built to filter the comments and screenshots by artifact, scenario, and cluster. Figure 10 shows the UI for searching the rater screenshots and corresponding comments. After selecting an application, scenario, and cluster, a result set appears that includes all relevant screenshots from the TEDS assessments. This tool greatly expedited the process of locating screenshots and comments for the feature improvement specifications; with a quick search, it presented the combined relevant information.

Figure 10. Access Form for Filtering and Searching Ratings, Comments, and Screenshots



In Figure 10, the Player Information scenario and the Navigation and Findability cluster were selected for the Real Madrid application. If no value had been selected in one of the lists, no filtering would occur for that list. For example, if no cluster had been selected, but Real Madrid and Player Information were selected, the tool would return all photos for the Real Madrid application for Player Information for all clusters. Figure 11 shows a results page of the screenshots and corresponding comments when the Real Madrid application, Player Information, and Navigation and Findability were selected.

Figure 11. Filtered Access Form Result with Ratings, Comments, and Screenshots



The search tool made it possible to quickly examine rating details and then brainstorm ideas for treatments for an artifact. The TEDS raters were not required to explain each rating or attach a screenshot in this study. Thus, some scenarios, artifacts, and cluster combinations do not include pictures and comments. The TEDS data set contained more comments than screenshots for the cluster ratings. Comments that explain the raters' rationale are critical for individuals trying to generate feature ideas, whereas the screenshots help demonstrate how the artifact used features or

UI components. Situations in which the rater provided a comment but not a screenshot for a specific scenario, artifact, and cluster combination, a screenshot could be taken by the researcher to assist in generating feature improvements. For this study, specific scenario and cluster combinations for the Sounders FC mobile application were compared with the highly rated mobile applications in the study to generate feature ideas for the Sounders FC application that could be A/B tested.

Focus Group Discussion: Assumptions and Constraints

Development organizations live within budgets and the priorities of the organization. It would be fruitless to suggest that an organization attempt to generate treatments on features for which they have no control over the development. For example, focus group participants discussed the ticketing features in this study. If this feature were implemented using an iframe for Ticketmaster, the Sounders FC would not be able to A/B test the feature.

Focus Group Discussion: A/B Testing

Although the Sounders chose the Player Information and Schedule, Results, and League scenarios for A/B experimentation, the soccer season was ending, which limited the time available to A/B test treatments with users. Therefore, the Sounders FC development team chose to prioritize the Player Information treatment improvement specification for A/B experimentation.

The questionnaire asked participants to discuss considerations of A/B experimentation. Three focus group participants had experience with A/B experimentation. The comments and corresponding discussion corroborated topics discussed in the A/B experimentation methods section of this dissertation. It was suggested that the technology stack used for the artifact development needs to support A/B testing. Since this study included web-based applications, artifacts could be A/B tested using Google Content Experiments or another web-based A/B stack. Each experiment required an Overall Evaluation Criteria (OEC) to determine which treatment or control wins in the experiment.

To help determine that an outcome occurs because of a specific change, well-designed experiments minimize the number of things changed between the control and the treatment. For example, if a UI modification to a treatment page requires that the page resides on a different server with higher capacity, which improves the latency (length of delay) of rendering pages, it can be difficult to determine whether the UI change caused the improvement, or if the quicker page rendering made the difference.

Organization of the Feature Improvement Treatment Specification

TEDS assessment data can be pivoted by scenario or cluster. Although using either pivot provides insight, a scenario-based TEDS analysis matches traditional functional feature specifications, which describe the intended functionality of an application or feature. The focus group determined that the treatment specifications should also provide the functional descriptions of the application and be scenario-based. A scenario-based functional specification could be created for each scenario that an organization sought to address using this method.

In this study, the Sounders would be provided a feature improvement specification for Player Information and another for Schedule, Results, and League. The information in the specification is organized by cluster, and includes a detailed discussion of the artifacts that performed well and the screenshot examples. In Appendix G, the Sounders FC Player Information Feature Improvement Treatment Specification includes a summary table of the TEDS ratings for each cluster in Player Information. From this table, it is easy to identify the higher-ranked artifacts for each TEDS cluster. The final section in the specification shows a summary of the recommended treatments for the Sounders FC mobile application for Player Information, pulled from the previous 12 cluster sections.

Detailed Analysis of Each Cluster

A detailed analysis of each cluster includes three sections:

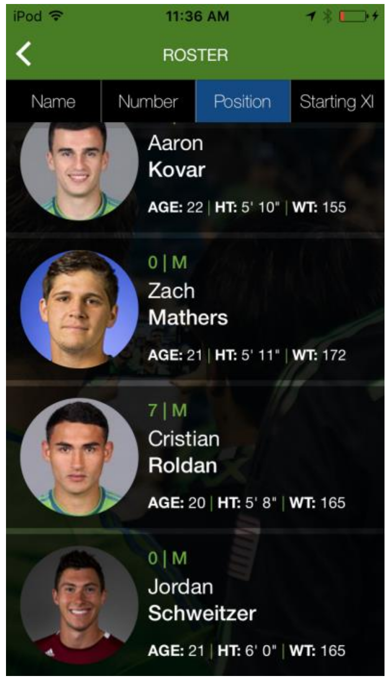
- Definition for each cluster
- Evaluations and screenshots from highly ranked artifacts for the cluster
- Sounder FC treatment ideas generated from analyzing the evaluations, comments, and screenshots

The initial process of building the specification is mechanical and iterative. The definition of each TEDS cluster is included in each cluster section as a reference for designers who are not familiar with the TEDS framework. The TEDS scenario-based ranking table is included for each cluster so that the designer can identify the artifacts that performed well for a given scenario or cluster. For example, the *Ratings – Navigation and Findability* table can be found in Appendix G, p. 6. Using the TEDS Rating, Comments, and Screenshot Search Tool, the designer locates the comments and screenshots from the high-performing artifacts. These screenshots and comments are pasted into the section of the specification for each TEDS cluster. This gives the designer an organized representation of information to help brainstorm treatment ideas to improve the performance of the artifact. This is the non-mechanical or creative part of building the specification. Treatment ideas are written in the specification for each cluster section below the screenshots and comments. The final section of the document includes the list of ideas that were generated for each cluster. The designer then has the full list of treatment ideas for review and prioritization.

Player Information A/B Treatments

The two main pages to consider for Player Information are the roster page and the player card page. Figure 11 shows how the roster page looked at the beginning of the TEDS assessment phase, before any feature treatments.

Figure 11. Roster page at beginning of the TEDS assessment phase

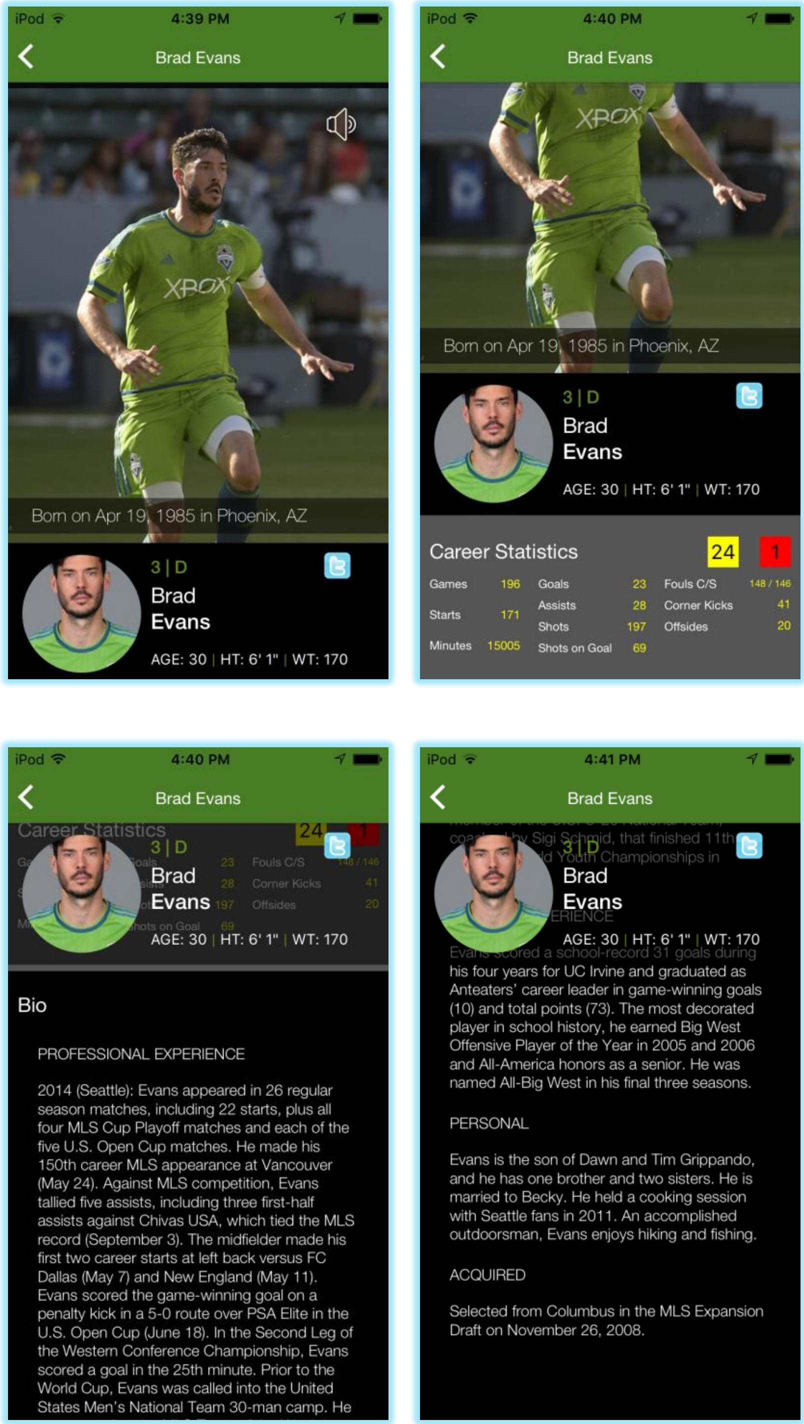


The sort order of the players in the roster was determined by the filters at the top of the screen (Name, Number, Position and Starting XI). Users could scroll through the full roster by swiping vertically on the screen, which scrolled the list of players up and down on the page. The Starting XI selection showed the players who were to start the upcoming game. This listing appeared a few hours before a game kickoff and for a day or so after the game. However, selecting this option at other times would show a blank page with no players.

As shown in Figure 12, during the first phase of TEDS assessments, each Sounders FC player card consisted of one page with four sections: a large action photo at the top, a smaller headshot with text about demographic information (age, height, weight, number, and position); a descriptive

Career Statistics section, and a Bio section with one or more pages of static text.

Figure 12. Sounders FC Player Card at beginning of the TEDS assessment phase



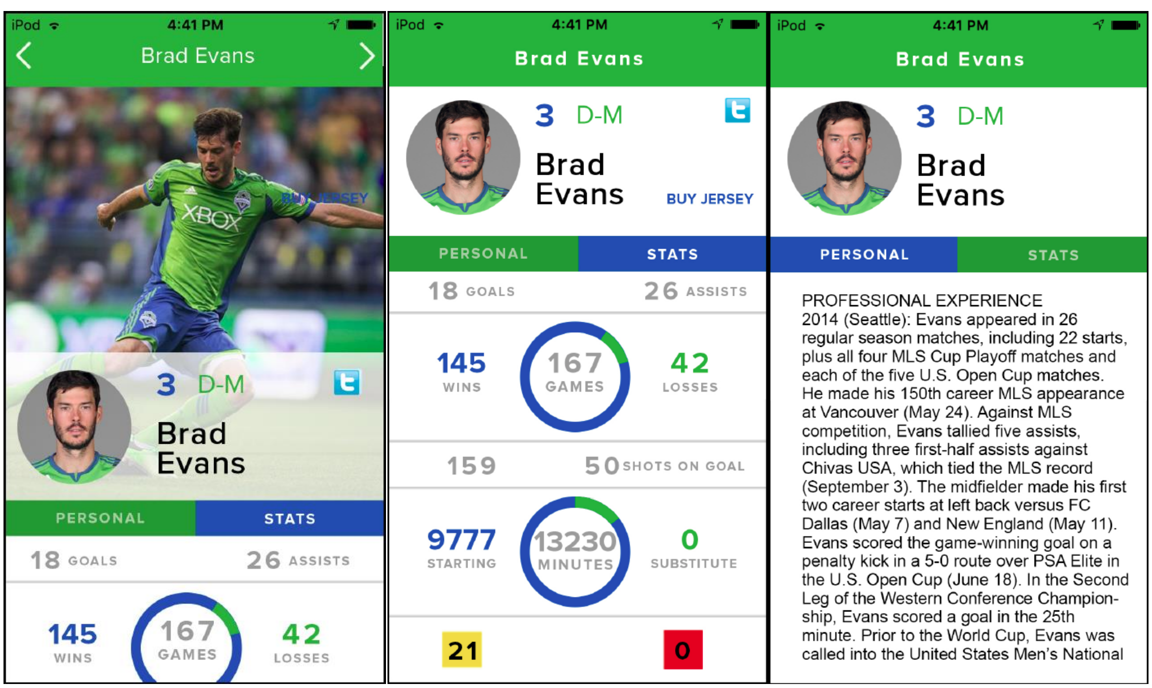
During the Treatment Development Phase, four treatment suggestions emerged and were added to the Player Information Feature Improvement Specification (See Appendix G, page 53):

- Add a tab that opens career statistics on a new page
- Add a large photo and horizontal scrolling to the roster page
- Show player statistics in a graph
- Link to the online store to buy a jersey

Add a tab that opens career statistics on a new page

The existing player card consisted of one scrolling page that felt overcrowded with information. The suggestion was to move the statistics to a page on a new tab. Based on the rater feedback, a tab is more discoverable than swiping; tabs are familiar because they are commonly used in other applications. Figure 13 shows screenshots of the design mockups that were presented to the Sounders FC development team.

Figure 13. Design mockups with career statistics on a new page

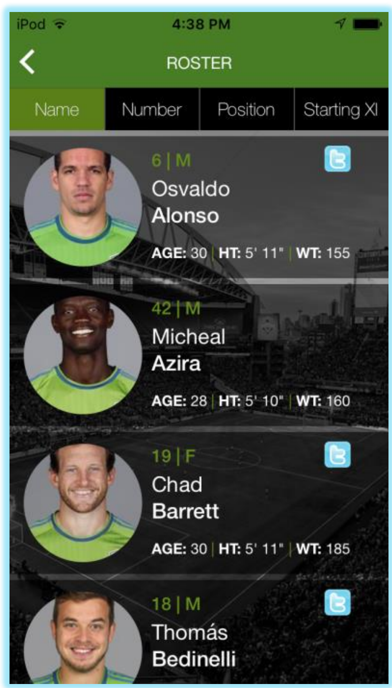


Tapping PERSONAL showed the Bio section; tapping STATS showed the career statistics.

Add a large photo and horizontal scrolling to the roster page

The existing Sounders FC roster page showed the players enumerated vertically with about four players shown at a time. The headshots were small, with little information about each player (age, height, weight). Figure 14 shows a screenshot of the Sounders FC roster page before any feature treatments.

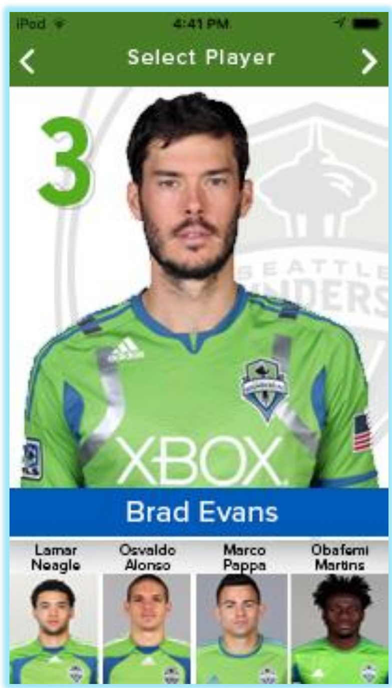
Figure 14. Existing Sounders FC roster page



In one of the highly rated iOS applications in this study, a large picture of a default player appears, along with a horizontal navigation at the bottom. This application used very little text to describe the player. But the page stood out with a large, clear photo of the player. Navigation to other players was easy to see, with smaller player photos appearing at the bottom of the page. By swiping on the roster at the bottom, the user can quickly scroll through the list of team members.

A proposed template for the roster page, shown in Figure 15, was presented to the Sounders FC development team. Using the right chevron at the top of the screen, a user could scroll to the next player in the roster, which would replace the large photo with the photo of that player and also update the player images at the bottom of the screen. Tapping a smaller photo at the bottom of the page replaces the larger photo with that player's photo. For example, if a user tapped the photo of Marco Pappa, his image would replace the image of Brad Evans, and an image of another player would appear in the roster at the bottom. To see the Player Card for the player in the large photo, the user would simply tap the larger photo.

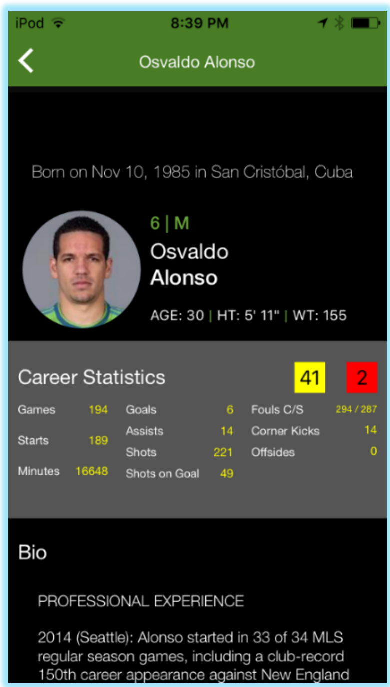
Figure 15. Design mockup of the roster page with a large photo and horizontal scrolling



Show player statistics in a graph

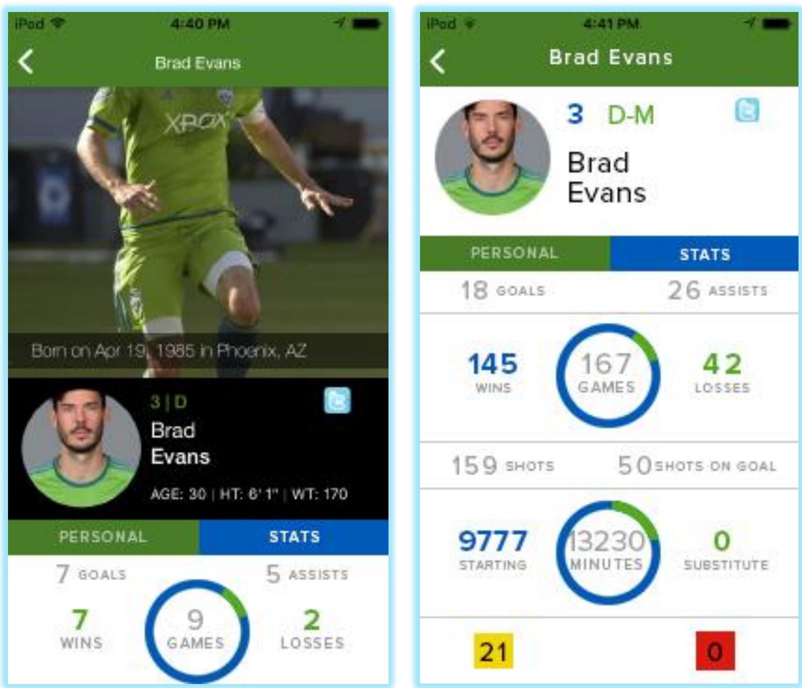
As shown in Figure 16, the existing Sounders FC player statistics appeared as static text on a grey box on the player card. Although the number of statistics was average compared to other soccer iOS apps, others applications display charts of the player statistics, which make the statistics easy to see at a glance and make the page more attractive.

Figure 16. Existing Sounders FC player statistics



To present player statistics in a graph instead, the Sounders FC development team was provided the design mockups shown in Figure 17. The circle graphs suggested in the mockup match graphs used by other applications in the study and by the Sounders FC on their team statistics page.

Figure 17. Design mockups for player statistics



Link to the online store to buy a jersey

The existing player card (shown in Figure 12) did not provide a link to the store to buy a jersey. This was a missed opportunity to drive revenue for the Sounders FC and also to provide an easy way for users to buy a jersey from the store. Users had to find the link to the team store and then search for a jersey for that player. Figure 18 shows the design mockup with a link to the store.

Figure 18. Design mockup with a link to the online store to buy a jersey

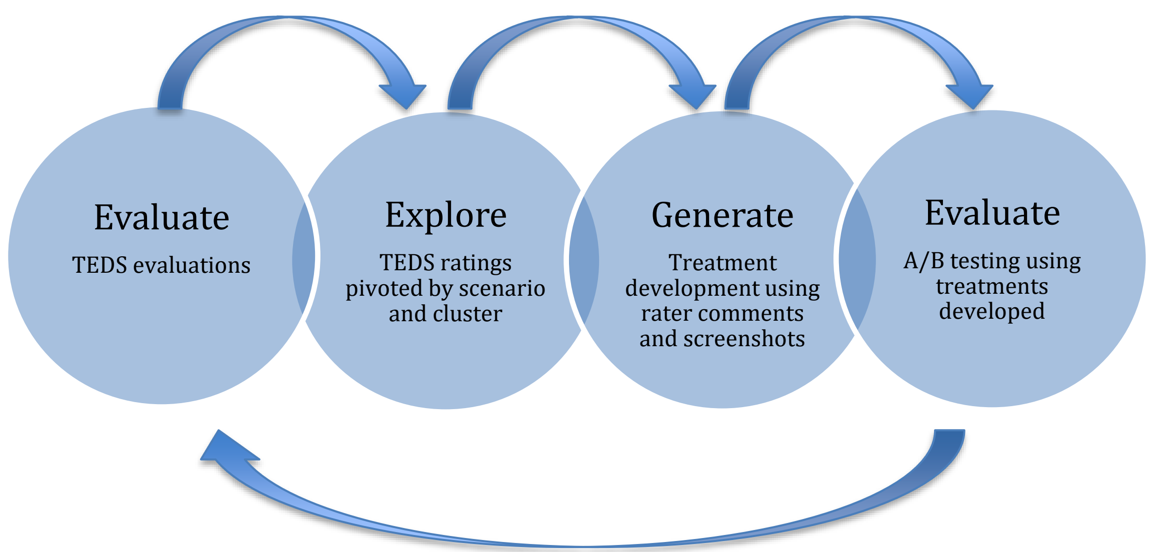
The following paragraph is a summary of the RQ2 findings.

RQ2 Findings Summary

The Delphi focus group method suggested using the Carnegie Mellon design education method to describe the integration of the TEDS method with the treatment design and A/B testing phases. The Carnegie Mellon model corresponds well to the approach in this dissertation because of the iterative nature of its design phases.

This study started with the TEDS evaluations. This led to the exploration phase, which used different views of the ratings to help rank the performance of the information artifacts in the study. The rankings, in turn, informed the generation phase in which the treatments specifications were developed using the screenshots and comments from the raters. These specifications were then used to develop treatments which were quantitatively evaluated in the A/B test phase. After A/B testing was completed and the successful treatments were deployed, the whole process could begin again with a new TEDS assessment. Figure 20 illustrates this iterative process.

Figure 20. Iterative Process of Evaluation, Exploration, and Generation Used in this Study



This study demonstrated the value of using TEDS rankings to identify the top-performing information artifacts for each scenario and TEDS cluster value. Then, based on the rankings, the screenshots and comments of the top-performing artifacts were organized in a feature improvement specification according to the TEDS clusters. This way, a designer could identify design patterns and best practices, and also use the TEDS evaluations to pivot and rank the data based on individual cluster ratings across all scenarios. This makes it easier to identify artifacts that perform well for a specific TEDS cluster value across the scenarios in the study. The two feature improvement specifications generated in this study were the Sounders FC Player Information Feature Improvement Treatment Specification, and the Sounders FC Schedule, Results, and League Feature Improvement Treatment Specification.

Addressing Research Question 3 (RQ3)

How can the number of values used in a TEDS assessment be reduced for evaluating an information artifact in a given context in order to focus on values that are most pertinent to users of the information artifact? Although this question was included in the dissertation proposal, the committee did not recommend investigating the pertinence of TEDS values for a given context because of the large scope of RQ1 and RQ2. From a practical perspective, the evaluations were a considerable task for each rater: This study included TEDS evaluations of eight scenarios of action/interaction for each of the 10 iOS mobile applications. In addition, all 40 sub-criteria were used to evaluate the FC Barcelona and Seattle Seahawks iOS mobile apps, and the 12 clusters were used to evaluate the remaining nine iOS mobile apps.

Phase 1 User TEDS Assessments

In Phase 1, during the TEDS procedure setup, as the expert raters were performing the inter-rater reconciliation of the anchor artifacts, they recognized that there would be a benefit of only rating the high-relevance sub-criteria. This would focus the evaluations on the meaningful sub-criteria and reduce having to rate non-relevant ones. However, they also recognized from their evaluations that there were patterns in a rating for many of the sub-criteria in a given scenario. The sub-criteria ratings tended to group together. This led to the insight of clustering related sub-criteria. Clustering allowed the raters to consolidate related sub-criteria to correspond with what they were noticing in the correlations among sub-criteria in their traditional 40 sub-criteria evaluations, without losing the ability to re-expand the cluster when desired.

TEDS evaluations that include all 40 sub-criteria ratings can be compared with TEDS evaluations that include 12 clusters by using the average rating of the sub-criteria included in each cluster for the 40 sub-criteria ratings. Although this method allows for easy comparison between the two methods, there are issues to consider when sub-criteria ratings are averaged if the different clusters for a given artifact are being compared. For example, the number of sub-criteria in the clusters varies. There are two sub-criteria in the confidence cluster (security, safety), and 6 sub-criteria in the customization cluster (flexibility, simplicity, trust, individualization, localization, privacy). It would be ill-advised to blindly compare the average confidence and customization cluster ratings, given that there are three times as many sub-criteria in the customization cluster. Further, clustering does not take into consideration the relevance of the individual sub-criteria for a given context. It is possible that one criterion is dramatically more important for an information artifact in a particular scenario.

The following paragraph is a summary of the RQ3 findings.

RQ3 Findings Summary

Although not anticipated during dissertation proposal time, in the course of the methodological development in this study, it became evident that the scalability of the TEDS assessment method not only allowed for but rather required a consolidation of assessment sub-criteria. As a consequence, a systematic sub-category clustering mechanism was developed, which maintained the consistency of the method and the integrity of the evaluative TEDS approach.

This study did not investigate a method of identifying which TEDS values were more or less pertinent for an information artifact in a given context. This remains a worthwhile research area. Analyzing ratings and comments from completed TEDS studies could provide insights. This study demonstrated that TEDS clusters could be employed to reduce the number of individual ratings in an evaluation while preserving the fine-grained ability of TEDS, because clusters could be re-expanded when desired for more precision. The clusters were used for both user and expert evaluations. Layman definitions of each cluster were generated for the user surveys.

Addressing Research Question 4 (RQ4)

Research Question 4 (RQ4): How effective is data from TEDS assessments gathered using surveys from a purposive sampling, in providing effective data to be used in combination with the expert ratings data provided in the TEDS method?

Similar to RQ3, the committee did not require this question to be included in the dissertation because of the large scope of RQ1 and RQ2. However, there was a benefit of having user-generated TEDS ratings for comparison between Phase 1 and Phase 4. (User ratings are presented for Player Information and Schedule, Results, and League in the Findings section for RQ1.) First, since all user raters were unique between Phase 1 and Phase 4, they were not subject to rater bias that could exist with the expert ratings.. Second, the development of a web-based TEDS user rater survey would provide data that could direct future research and enhancements to the method. Thus, the TEDS user survey and method was introduced and the comparison of expert and user ratings are presented in the following sections.

Phase 1 User TEDS Assessments

The Phase 1 TEDS user surveys were active between May 19, 2016 and June 22, 2016. During those five weeks, 73 Player Information assessments and 40 Schedule, Results, and League assessments were collected. Surveys from users who did not meet the purposive sample requirements (rarely or never watched soccer, or were not a fan or follower of a team) or those who completed fewer than five cluster ratings were removed. After paring, 56 Player Information assessment surveys and 31 Schedule, Results, and League surveys were included in the data set. (See Appendix K for the full list of Phase 1 user TEDS ratings for Player Information and Schedule, Results, and League.) Table 20 shows the Player Information chart for the Sounders FC application, with detailed cluster ratings by users and experts, and their correlation.

Table 20. Phase 1: Sounders FC Mobile Application User and Expert Ratings for Player Information

Phase 1: Sounders FC Mobile Application User and Expert Ratings for Player Information						
Cluster	Users	Experts	Difference	Adjusted Users	Experts	Difference
Navigation and Findability	3.46	4.00	-0.54	3.15	4.00	-0.85
Structure	3.59	3.60	-0.01	3.27	3.60	-0.33
Identity	3.96	3.50	0.46	3.65	3.50	0.15
Substantiality	3.68	3.60	0.08	3.36	3.60	-0.24
Completeness	3.91	3.20	0.71	3.59	3.20	0.39
Trustworthiness	4.05	3.70	0.35	3.74	3.70	0.04
Interaction	3.34	3.30	0.04	3.02	3.30	-0.28
Customization	3.50	2.60	0.90	3.18	2.60	0.58
Savings	3.63	3.00	0.63	3.32	3.00	0.32
Confidence	3.80	3.30	0.50	3.49	3.30	0.19
Attractiveness	3.61	3.20	0.41	3.29	3.20	0.09
Enjoyment	3.45	3.20	0.25	3.13	3.20	-0.07
Average	3.67	3.35	0.32	3.35	3.35	0.00

The values in Table 20 are conditionally formatted from dark green to yellow. Dark green indicates the highest values; yellow indicates the lowest values. Where colors for the user and expert ratings in a cluster are similar, users and experts rated the cluster similarly on average relative to the other clusters in this scenario. The comparison table on the left shows the raw averages for users and experts. Navigation and Findability was the only cluster that had a lower average user rating than the average expert rating (users: 3.46, experts 4.00). On average, the user ratings were 0.32 higher than the expert ratings. In the comparison table on the right, each user rating is adjusted by subtracting 0.32 from the average user cluster rating. This conversion was added to make it easier to compare the normalized cluster ratings of the two sources.

Table 21 compares user and expert ratings for Schedule, Results, and League. The only two clusters for this scenario that had a lower average user rating than the average expert rating were

Navigation and Findability (users: 3.68, experts: 3.90) and Structure (users: 3.45, experts 3.60). The average difference between the user and expert ratings was again 0.32. A subsequent study that includes a larger sample size of applications across more scenarios will help identify trends. For this study, the users and the expert ratings provided a baseline for comparison later in Phase 4, after the Phase 3 A/B testing with the developed treatments was completed, and the winning treatments had been deployed.

Table 21. Phase 1: Sounders FC Mobile Application User and Expert Ratings for Schedule, Results, and League

Phase 1: Sounders FC Mobile App User and Expert Ratings for Schedule, Results, and League						
Cluster	Users	Experts	Difference	Adjusted Users	Experts	Difference
Navigation and Findability	3.68	3.90	-0.22	3.36	3.90	-0.54
Structure	3.45	3.60	-0.15	3.13	3.60	-0.47
Identity	3.68	3.60	0.08	3.36	3.60	-0.24
Substantiality	3.68	3.30	0.38	3.36	3.30	0.06
Completeness	3.71	3.10	0.61	3.39	3.10	0.29
Trustworthiness	3.90	3.60	0.30	3.58	3.60	-0.02
Interaction	3.17	2.30	0.87	2.85	2.30	0.55
Customization	3.39	2.40	0.99	3.07	2.40	0.67
Savings	3.48	3.40	0.08	3.16	3.40	-0.24
Confidence	3.81	3.00	0.81	3.49	3.00	0.49
Attractiveness	3.48	3.40	0.08	3.16	3.40	-0.24
Enjoyment	3.42	3.40	0.02	3.10	3.40	-0.30
Average	3.57	3.25	0.32	3.25	3.25	0.00

Phase 4: Second TEDS Assessment for the Sounders FC iOS Application

On November 5, 2016, the Sounders FC development team deployed an update to their iOS application to activate both user treatments from the A/B experiments to all users. This voluntary update was visible when users logged on to the Apple App store. Users in the control group continued to see the control version of the roster page and the text-based statistics module on the player card until they installed the new version from the app store.

To help ensure that users installed the new version, the material to recruit survey takers emphasized the requirement to first update their application to the latest version before running the survey. Similarly, all expert raters were told to install the new version before doing their Phase 4 assessments. For the survey takers recruited on Facebook and Instagram, no mechanism was in place to confirm they had installed the latest version. However, it seems unlikely that most users who were offering to take the survey would not follow the instructions. There were no screenshots or comments in the Facebook or Instagram survey data that referred to the control versions of the roster page or to the control text-based statistics module on the player card.

In the second TEDS assessment conducted in Phase 4, the same expert raters who were in the Phase 1 TEDS assessment of the Sounders FC application were contacted by email and asked to assess the updated Sounders FC application for all eight scenarios used in Phase 1. To give raters context while completing the evaluations, they referred to anchor data information from the Phase 1 assessments. One strength of the traditional TEDS method is raters can compare and consider evaluations from other artifacts. As experts become familiar with ratings of other artifacts, there tends to be smaller deviation between their evaluations. (See Appendix L for Phase 4 full expert ratings for the Player Information and Schedule Results and League scenarios. See Appendix M for Phase 1 full expert ratings for these two scenarios.)

The Phase 4 TEDS assessments of users recruited from Facebook and Instagram ran for 30 days from November 5-December 5, 2016. The recruitment used the identical advertisement from Phase 1. Users who clicked the ad were taken to a webpage that described the study and provided links to the Player Information and Schedule, Results, and League surveys. The only update to the webpage for the second assessments in Phase 4 was instructions that the survey taker install the new version of the Sounders FC iOS application from the Apple App Store.

Eighty-one Player Information and 47 Schedule, Results, and League assessments were collected. After removing users who did not meet the purpose sample requirements or who completed fewer than five cluster ratings, 77 Player Information and 42 Schedule, Results, and League surveys were included in the data set. Table 26 shows a comparison of user and expert ratings in Phase 4 for Player Information.

Table 26. Phase 4: Sounders FC Mobile Application User and Expert Ratings for Player Information

Phase 4: Sounders FC Mobile Application User and Expert Ratings for Player Information						
Cluster	Users	Experts	Difference	Adjusted Users	Experts	Difference
Navigation and Findability	3.71	4.10	-0.39	3.50	4.10	-0.60
Structure	3.84	4.15	-0.31	3.64	4.15	-0.51
Identity	3.97	3.70	0.27	3.77	3.70	0.07
Substantiality	3.67	3.90	-0.23	3.47	3.90	-0.43
Completeness	3.87	3.60	0.27	3.66	3.60	0.06
Trustworthiness	4.15	3.80	0.35	3.94	3.80	0.14
Interaction	3.26	3.50	-0.24	3.06	3.50	-0.44
Customization	3.48	2.70	0.78	3.28	2.70	0.58
Savings	3.68	3.10	0.58	3.48	3.10	0.38
Confidence	4.11	3.40	0.71	3.91	3.40	0.51
Attractiveness	4.10	3.55	0.55	3.89	3.55	0.34
Enjoyment	3.71	3.60	0.11	3.50	3.60	-0.10
Average	3.80	3.59	0.20	3.59	3.59	0.00

For the Player Information scenario in Phase 4, users rated the average Sounders FC iOS mobile app 0.20 higher than the expert raters. In Phase 1, the average difference was 0.32. Table 27 shows a comparison of user and expert ratings in Phase 4 for Schedule, Results, and League.

Table 27. Phase 4: Sounders FC Mobile Application User and Expert Ratings for Schedule, Results, and League

Phase 4: Sounders FC Mobile App User and Expert Ratings for Schedule, Results, and League						
Cluster	Users	Experts	Difference	Adjusted Users	Experts	Difference
Navigation and Findability	3.93	4.00	-0.07	3.63	4.00	-0.37
Structure	3.81	3.80	0.01	3.51	3.80	-0.29
Identity	3.93	4.00	-0.07	3.63	4.00	-0.37
Substantiality	3.86	3.50	0.36	3.56	3.50	0.06
Completeness	3.95	4.10	-0.15	3.66	4.10	-0.44
Trustworthiness	4.17	4.00	0.17	3.87	4.00	-0.13
Interaction	3.33	3.10	0.23	3.03	3.10	-0.07
Customization	3.61	2.50	1.11	3.31	2.50	0.81
Savings	3.95	3.50	0.45	3.65	3.50	0.15
Confidence	4.27	3.10	1.17	3.97	3.10	0.87
Attractiveness	4.10	3.75	0.35	3.80	3.75	0.05
Enjoyment	3.85	3.85	0.00	3.55	3.85	-0.30
Average	3.90	3.60	0.30	3.60	3.60	0.00

For the Schedule, Results and League scenario in Phase 4, the average user rating for the Sounders FC iOS mobile app was 0.30 higher than the expert raters. In Phase 1, the average difference was 0.32. On average, users rated the application higher than expert raters. This average difference was not consistent for each cluster, but a correlation did exist between clusters when comparing different phases. In Table 28, comparison of the Phase 1 and Phase 4 user and expert cluster assessment data sets, a strong correlation shows up in Customization, Savings, Confidence, Attractiveness, and Schedule, Results, and League.

Table 28. Comparison of Phase 1 and Phase 4 user and expert cluster ratings for Player Information

Phase 1: Sounders FC App - Player Info				**Phase 4: Sounders FC App - Player Info**			
	Users	Experts	Difference		Users	Experts	Difference
Navigation and Findability	3.46	4.00	-0.54	Navigation and Findability	3.71	4.10	-0.39
Structure	3.59	3.60	-0.01	Structure	3.84	4.15	-0.31
Identity	3.96	3.50	0.46	Identity	3.97	3.70	0.27
Substantiality	3.68	3.60	0.08	Substantiality	3.67	3.90	-0.23
Completeness	3.91	3.20	0.71	Completeness	3.87	3.60	0.27
Trustworthiness	4.05	3.70	0.35	Trustworthiness	4.15	3.80	0.35
Interaction	3.34	3.30	0.04	Interaction	3.26	3.50	-0.24
Customization	3.50	2.60	0.90	Customization	3.48	2.70	0.78
Savings	3.63	3.00	0.63	Savings	3.68	3.10	0.58
Confidence	3.80	3.30	0.50	Confidence	4.11	3.40	0.71
Attractiveness	3.61	3.20	0.41	Attractiveness	4.10	3.55	0.55
Enjoyment	3.45	3.20	0.25	Enjoyment	3.71	3.60	0.11
Average	3.67	3.35	0.32	**Average**	3.80	3.59	0.20

Table 29. Comparison of Phase 1 and Phase 4 user and expert cluster ratings for Schedule, Results, and League

Phase 1: Sounders FC App - Schedule, Results, and League			
	Users	Experts	Difference
Navigation and Findability	3.68	3.90	-0.22
Structure	3.45	3.60	-0.15
Identity	3.68	3.60	0.08
Substantiality	3.68	3.30	0.38
Completeness	3.71	3.10	0.61
Trustworthiness	3.90	3.60	0.30
Interaction	3.17	2.30	0.87
Customization	3.39	2.40	0.99
Savings	3.48	3.40	0.08
Confidence	3.81	3.00	0.81
Attractiveness	3.48	3.40	0.08
Enjoyment	3.42	3.40	0.02
Average	3.57	3.25	0.32

Phase 4: Sounders FC App - Schedule, Results, and League			
	Users	Experts	Difference
Navigation and Findability	3.93	4.00	-0.07
Structure	3.81	3.80	0.01
Identity	3.93	4.00	-0.07
Substantiality	3.86	3.50	0.36
Completeness	3.95	4.10	-0.15
Trustworthiness	4.17	4.00	0.17
Interaction	3.33	3.10	0.23
Customization	3.61	2.50	1.11
Savings	3.95	3.50	0.45
Confidence	4.27	3.10	1.17
Attractiveness	4.10	3.75	0.35
Enjoyment	3.85	3.85	0.00
Average	3.90	3.60	0.30

Several things could cause this correlation. For example, each expert rater received TEDS training, including an inter-rater reliability discussion about the anchor artifacts. Experts also had more insight into the definitions of each cluster. The survey instrument could also contribute to misunderstanding for users if the questions were not written clearly. These are all questions that a subsequent study should investigate.

The following paragraph is a summary of the RQ4 findings.

RQ4 Findings Summary

Though this was not anticipated by the supervisory committee, it was found critical to introduce and develop the purposeful sample method for gathering user TEDS evaluations used in this study. Comparing the results showed some correlation between user and expert ratings in the corresponding phases for the two scenarios. On average, users rated the artifacts higher on the Likert scale than the experts rated them. The results raised some issues regarding whether the questions for the individual clusters were clear to the user raters. This will be investigated in future research.

Chapter 5

DISCUSSION

Treatment creation using the TEDS assessment artifacts

As shown by the findings of this study, the use of the TEDS methodology with the addition of a treatment creation phase provides a systematic way for an organization to act on the findings from a TEDS study. It could be used as a structured competitive analysis methodology, with ratings and evaluations for a set of similar information artifacts. It could also include recommendations to improve the competitiveness of the information artifacts in the assessments.

During the treatment creation phase, the designer uses comments and screenshots (in addition to the Likert ratings) from assessments. The Likert ratings allow artifacts to be ranked by scenario or cluster. But if there are no comments or screenshots, the interpretation is missing. This is an area of potential weakness in the TEDS procedure. It is optional for raters to add comments and screenshots with an artifact evaluation. In this study, the expert assessments included specific comments and related screenshots that were used to generate the treatment improvement specifications. For comparison purposes, in the initial TEDS evaluation phase, five experts participated, and provided 20 comments and 9 screenshots when assessing the Player Information scenario. Fifty-six users evaluated Player Information and provided 50 comments and three screenshots. Many of the user comments were generic and not actionable.

Since this study used the expert assessment data to rank the scenarios and generate treatments, the quality of the data for Phase 2 was sufficient. In the future, work should be done to gather better quality comments and screenshots from users. The user assessment tool provides a method to attach screenshots along with help content to assist raters through the process. But there might be better ways to capture the comments and screenshots. One possibility would be to show a short video on how to attach screenshots before the survey. Another possibility is to have the tool take screenshots of the application for the given scenario, and then ask the user to select an existing screenshot that they can comment on. Or, the survey tool could include a voice recording feature so users can explain their ratings verbally. Encouraging users to verbalize their ratings could garner richer, more actionable feedback.

Although generating the treatment improvement spec was tedious for the first scenario, it served as a template for the second treatment specification.

Research embedded in a live development organization

Unlike a laboratory experiment in which a researcher can neutralize outside influences, this project was situated in a live development project. Projects like the Sounders FC iOS mobile app evolve organically over time. Although this presented some challenges in interpreting the magnitude of

TEDS ratings changes from Phase 1 to Phase 4, the benefits of embedding a dissertation in a real-world project are many. First, this project demonstrated several adjustments that can be made real-time during the project so that it could move forward without reducing the benefit of the method being applied. There are many examples of this, but two that stand out are the decision to use Flurry rather than Google Content Experiments as the A/B testing platform, and the process that the Sounders used to determine which scenarios and treatments they experimented with.

Although the Sounders never explicitly exposed the amount of development resources they were willing to provide this project, they made it clear that their existing backlog of development items had priority.

Every project has three variables to work with: time, resources, and scope. The development resources were limited. In fact, only one dedicated developer was working on the frontend of the iOS application. The Sounders were not going to increase resources for this project. The time was fixed and tied to the end of the MLS season. Delaying the project would have meant a minimum four-month hiatus from A/B testing the experiments while the Sounders were not in season. Also, the Sounders would likely have new priorities for their mobile application in the off season. So, the scope of this project had to be adjusted.

Notably, the Sounders FC development team used the TEDS rankings as only one input to determine which scenarios to address in Phase 2 and Phase 3. Their appraisal of the relative importance of the scenarios was the other important input. In the Phase 1 TEDS assessment, the Sounders FC iOS application had ranked 6th out of 10 compared to the other artifacts in the Player Information and Schedule, Results, and League scenarios. The Sounders positioned these scenarios ahead of Social Media Integration, even though the application ranked worse (8 out of 10) for that scenario. It was more important for them to address slightly lagging features in high value scenarios ahead of less important features that were below average in lower value scenarios. This demonstrates that prioritization is a multi-variable consideration in a development organization. Although more experiments would have been interesting to include, it was not necessary. Two significant treatments were developed, A/B tested, and deployed in a development project with a fixed development budget and a fixed release date.

A/B Experimentation Phase

Many A/B testing platforms could have been used for this dissertation. However, the constraints of a real-world development project such as the Sounders FC iOS mobile application help determine the technology choice. The choice to use Flurry is a good example. The Sounders FC development team was already using Flurry, so they could perform basic A/B testing quickly without much additional development effort. Using an unfamiliar platform might have delayed the project and added risk to completing the final phases.

It is possible that the *Player Detail – Visible* event was not a good OEC for this experiment. This event tracked how many times the user navigated to the player card. The A/B experiments used this metric to indicate that the more the user navigated on average to the player card, the more delighted they were with the feature. Yet, it could be that more clicks could indicate that the user is confused or not finding the information they are looking for. Thus, the average session time for a user interacting with the feature might be a better OEC.

However, using a metric related to session time is problematic in that it is unclear whether users would spend more time on average looking at graphical representations or player statistics, or less time. They might find the information they need quicker and be satisfied, or they might be enticed to spend more time looking at statistics if they are presented in a more appealing fashion. At the time of this study, Flurry lacked many good options for the OEC, and there was limited time to work with the Sounders FC development team to instrument additional metrics. Therefore, average number of Player Detail – Visible events per user was determined to be the best option because it was simple to track and it seemed most likely that a higher average number of navigations would indicate more positive user engagement.

Users were assigned to either the control or the treatment for the two experiments. This means that users in the control group had the control version of the application for both experiments. Since Flurry Analytics lacks built-in A/B experimentation support, the assignment to treatment or control had to be implemented by the Sounders FC development team. Given their time constraints, it was not feasible to build a more sophisticated assignment mechanism in time to complete the study before the Sounders 2016 season ended. Completing the A/B testing during the season was essential for experimenting with users of actively using the application as intended. The Flurry data metrics show significantly different usage patterns by users with the application when the team is not playing in season.

As is often the case in a development organization, there was turnover in the Sounders FC development team just as the experiments were about to begin. The main developer who had participated from the beginning of this project left to take a position in a startup. A new developer on the Sounders FC development team needed to be brought up to speed. He was introduced to this study, but understandably, this was only one of his tasks, which slowed progress significantly.

Fortunately, the Sounders soccer team won their first two rounds of the playoffs, which extended the season. This allowed the experiments to continue running for an extra two weeks

How time affects IT assessments

Time has a complicating effect on TEDS assessments longitudinally. Online artifacts evolve, just as the Internet ecosystem in which the artifacts exist evolves. The TEDS assessments are useful for a limited period, as evidenced by comparing the Phase 1 to Phase 4 ratings for the Sounders FC iOS app across the eight scenarios. The Phase 1 expert ratings on the Beta version of the app took place in November 2015, a few months before the 2016 season began. The Phase 4 expert ratings took place in November 2016 before the end of the 2016 MLS season. There was only one year between the evaluations, yet the average difference in the expert scenario ratings between

Phase 1 and Phase 4 was 0.67. It is not surprising that a Beta rating would be substantially different than the final released product rating, but even after removing the scenarios that were in Beta form during the Phase 1 assessments, the average difference in the expert ratings for the remaining scenarios was 0.22. Although this might be explained as merely a variance of expert ratings, the number of development changes by the Sounders FC development team, as indicated by their list of deployed feature items, between phases (see Appendix O) suggests that the application was improved.

The user ratings also changed significantly. The time difference between the start of the Phase 1 and start of the Phase 4 user assessments was five and half months (May 19 - Nov 5, 2016). For the two scenarios, the Player Information average rating improved 0.13 and the Schedule, Results and League average rating improved 0.33 for an average increase of 0.23 in Phase 4.

These changes exemplify the time criticality of IT assessments. Had all the artifacts in the study been reevaluated in Phase 4, change would be expected in all artifacts ratings because, like the Sounders application, information artifacts continue to evolve. Or, the Internet ecosystem might have changed during that period, with a new technology adopted by other artifacts, making users expect that behavior elsewhere and thus changing the rubric for evaluation. The organic nature of information artifacts and the Internet ecosystem underscores why a development methodology needs to be iterative to remain competitive.

In Phase 1, the expert raters were enrolled at the University of Washington and although they were busy, most completed their assessments within a few days. By the time the Phase 4 TEDS assessments were to be done, only one of the experts remained at the University of Washington. This made it was much more difficult for the experts who left the area to squeeze in time to complete the assessments, with new jobs in locations around the world. It took one month to complete the expert ratings in Phase 4.

A time criticality element to the expert assessments also emerged. To achieve consistency in their evaluations, expert raters leverage the anchor artifacts and their recent memories of how they rated other artifacts. In Phase 1, the experts each rated about two artifacts a week. Each week there was opportunity for inter-rater reliability discussion. As the weeks proceeded, the raters were very consistent among themselves because they had recent memories of how they had calibrated their ratings. However, in Phase 4, it seemed more difficult for the raters to recall the calibrations and the common understanding they had built collectively in Phase 1. The amount of time between assessments is understood to be a factor in inter-rater reliability in the literature (Hays & Revicki, 2005). Yet, this seems to be an issue only for expert evaluations, where there is typically a small number of raters. Using user evaluations, there is no need to do inter-rater reliability discussions, because discrepancies diminish the larger the sample size provided it is an unbiased sample of users doing the evaluations.

Transferability of this iterative process using a different assessment framework

Chapter 2, Literature Review, enumerates many useful assessment frameworks from academic literature. At the time of this writing, the TEDS framework was the most comprehensive taxonomy of any of the frameworks. Yet, this comprehensiveness was onerous when evaluating artifacts because of the 40 values to consider in the TEDS framework. This verbosity was the motivation for RQ3 of this study: to find a way to reduce the number of values considered for an artifact in a given context.

This study demonstrated that the TEDS sub-criteria could be clustered to efficiently reduce the number of ratings required for an individual evaluation, while retaining access to the fine-granularity of TEDS by re-expanding the cluster to the individual sub-criteria when desired. The Phase 1 anchor artifacts (FC Barcelona and Seattle Seahawk iOS mobile applications) were evaluated using all 40 TEDS sub-criteria, and their cluster ratings were calculated by averaging the sub-criteria within the corresponding clusters. This made it possible to compare the two anchor artifacts to the other eight artifacts in the study, which were assessed using the 12 clusters.

It could be argued that keeping the finer granularity sub-criteria while removing sub-criteria with low relevance from the framework for a given context would be even better. Within the TEDS method, this is something that should be considered for future research. However, if it is possible to use only a portion of the TEDS framework when assessing information artifacts, replacing the TEDS framework with a different academic framework is doable. In some contexts, the different framework's characteristics might be a great fit for evaluation.

Several recent examples in literature show that frameworks other than TEDS have been used to evaluate information artifacts. It would be possible to substitute the 12 clusters from the TEDS framework with the criteria from a different framework. However, TEDS remains the most robust framework, and the additional method to generate targeted feature improvements treatments using the ranked screenshots and comments the TEDS assessments, adds to the usefulness of TEDS.

TEDS Assessments and A/B testing

Although TEDS assessments and A/B tests are both evaluative and provide insight that helps prioritize feature treatments, they serve different purposes. During the TEDS evaluations, both the users and experts knew they were evaluating something. They were asked to rate an artifact in a specific context and to provide comments and screenshots to explain their opinions. In contrast, A/B tests gather user behavioral log data without the user's knowledge that they are participating in an experiment. This data mining of user behavior sometimes reflects a reality of user preference that the user does not consciously realize. However, that is a discussion for another day. TEDS remains a rigorous method to help prioritize feature improvements.

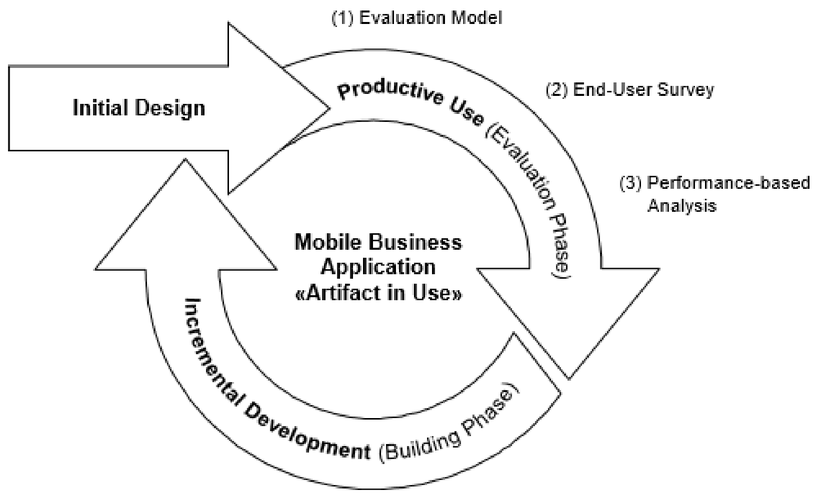
Focus group insights

Two additional topics were raised in the Delphi focus group about the creation of treatment improvement specifications. First, organizations have constraints that must be considered when making development investments. For example, many websites generate revenue through online advertisements. The TEDS evaluations might propose feature improvements that eliminate or dramatically reduce advertisements, which would not be possible for the organization. Second, the treatments being produced need to conform to the requirements of A/B testing. For example, A/B tests require that usage metrics are gathered. Features that are used off-line, when there is no mechanism for gathering and later synchronizing the off-line metrics, would not be candidates for A/B testing. Understanding the assumptions and constraints for the organization provides insight when generating treatments.

Recent Additions to the Literature with Relevance to the Study Findings and Future Application

Perhaps the most relevant recent contribution to the literature related to this dissertation is the use of a framework by Legner, et al. to evaluate information artifacts that feeds into an iterative design process (Legner et al., 2016). They adapted the D&M IS model and used it to gather user feedback through a survey. Just as in this dissertation, the assessment output enabled them to derive priorities and recommendations for their iterative design process in addition to generating feature treatments for the information artifacts evaluated. Interestingly, their study did not find support that the D&M Information Quality dimension was a motivator for consumer-oriented data services. Given the TEDS perspective, that context from a human-actor perspective is required to determine value, this finding can be explained that, in the context of their study, the attributes and characteristics related to information quality were not significant. Figure 19 demonstrates the iterative process they used (Legner et al., 2016).

Figure 19. Legner Diagram of Evaluation of Information Artifacts in an Iterative Design Process



Iterative Design Cycle for Mobile Business Applications

This diagram likely reflects what the Sounders FC development team did in going beyond using the TEDS assessments to evaluate and improve the Player Information and Schedules, Results, and League scenarios. The Sounders were provided the Phase 1 TEDS assessment results and rankings, which they were free to add to their development backlog. Assuming they used information from the assessments, which led to backlog work items, the Sounders did not formalize the process with written feature improvement specifications, nor use A/B testing as a quantitative evaluation tool to validate each feature's benefits before deploying the feature.

This dissertation is relevant to the Legner, et al. work. Using their end-user survey, they could follow the TEDS method with a set of similar information artifacts and integrate the gathering of screenshots and comments in user assessments. They could then optimize the treatment generation process in a similar fashion to what was done in Phase 2 of this dissertation. Finally, they could add A/B testing as a quantitative evaluation tool to validate the treatments before deploying them to all users.

Framework-based online surveys

This dissertation introduced an online instrument to gather TEDS evaluations from users. Several recent contributions regarding the application and theory of online surveys to measure information artifact quality can help TEDS researchers improve the online TEDS user instrument. Eapen and Chapman built a Likert scale survey (based on the D&M IS assessment survey) for users to assess the usability, productivity, and quality of mobile access to a healthcare tool called ClinicalConnect (Eapen & Chapman, 2015).

Shchiglik, et al. modeled an assessment questionnaire based on TAM, D&M, and other frameworks to evaluate mobile games (Shchiglik et al., 2016). Starting with 40 individual characteristics, they used a focus group and online survey to reduce the characteristics to 23, and then organized them into five dimensions: ease of use, ease of access to quality content, responsiveness, aesthetic appeal, and gaming experience. Although Shchiglik's paper does not include detail about the origin of each characteristic or the methodology, it provides an empirically validated framework and questionnaire that can be used to evaluate mobile games.

Hussain, et al. used the TAM model to evaluate user intention and acceptance of interactive mobile maps (Hussain et al., 2016). They developed a survey and used purposive sampling to recruit users at Universiti Utara Malaysia. The study examined users' perception of perceived usefulness, perceived ease of use, and perceived enjoyment as drivers of their complex interaction with maps on mobile devices. The analysis and findings showed that all three constructs have significant positive influence on user acceptance of interactive mobile maps. This study further validates the TAM model. Assila, et al. reviewed more than 20 usability questionnaires in the literature (Assila & Ezzedine, 2016) and analyzed more than 475 questions about the usability criteria used. This examination could provide insight about the TEDS user rating survey going forward.

Reflections on my own journey

Four years were spent (including the master's program) commuting to the UW campus for coursework before the general exam, dissertation proposal, and finally this dissertation. I've juggled full time work most of the time. The first four years of graduate coursework were grueling, but normal courses are structured. I kept on track because there was a syllabus, exams, and other deliverables that were mapped out in an orderly fashion. The challenge was making forward progress once the coursework was complete.

After completing the general exam, I changed roles at Microsoft, from a Test Manager position in MSN to a Program Manager position on the Bing A/B Experiment Platform team. It was a risky career move. I had been in the Test organization for more than 20 years. I went from managing a large team of software test engineers, to a being an individual contributor in a discipline that was new to me, with a completely new technology and different managers. I wanted to become an expert in A/B experimentation and this was the only position available on the best A/B experimentation team at Microsoft.

I saw the opportunity to use A/B experimentation alongside an assessment framework, and chose to jump. I was on the A/B team for a little more than two years. The change was positive because I learned A/B experimentation. I worked with very smart people on ground-breaking technology. But it was a very stressful time. Although I gained knowledge about A/B experimentation, the intensity of the job and the competitive nature of the team meant I had very little time to focus on my academic work, which prevented me from making any progress on the dissertation. On March 13 2014, I left Microsoft and took a position with fewer hours so I could make progress on this dissertation.

My initial plan for the dissertation was to use the Bing A/B experiment data from Microsoft to do empirical work on the TEDS attributes. I had planned to code existing experiments from Microsoft with the TEDS attributes and then observe the impact of these attributes across their huge collection of historical experiments. I had discussed this plan with my management team for several months prior to leaving the Microsoft and they agreed to let me do it provided it was accepted by the legal team and the senior executives in my division. I would provide Microsoft with all the data and insights found. Within a few weeks of leaving Microsoft, I began to work with the team to get my credentials.

As luck would have it, the week that my credential request was being processed, the article "Experimental evidence of massive-scale emotional contagion through social networks" (Kramer et al., 2014) by Kramer, Guillory, and Hancock, had just been published and was getting significant negative attention in the press. This article described negative A/B experiments that had been conducted using confidential user data. Although my proposed work was not related to this type of investigation, it is understandable why a highly-discussed article like this would evoke more scrutiny of my request for credentials to analyze confidential experimental data.

The management team spent time trying to understand what I proposed to do for my study. Their feedback was that it would be almost impossible to effectively code the TEDS attributes to the historical experiments without significant interaction with the experimenters. The experimenters who used the A/B platform had not consistently documented their experiments. The A/B team had found it difficult to determine details of historical experiments without significant help from the experimenter. Even with access to experimenters, it was not clear that they would remember details from many of the older experiments. This led the management team to believe that the likelihood of success of my original dissertation proposal was low. In addition, the timing of the Kramer article showed that there was some risk to Microsoft by providing credentials for academic research.

Although this setback was disappointing at the time, I appreciate the analysis that the Bing A/B Experiment Platform team did for me. Had I been given permission to the data, I likely would have run into the problems they foresaw. Instead, I began to both rescope and expand the scope of the dissertation proposal, which required substantial methodological effort through the four phases. The result is an applied dissertation that can be leveraged by other researchers. Certainly, there is opportunity for future research to do more empirical testing of the TEDS framework, but the work here shows great promise for both academia and industry.

Chapter 6

CONCLUSION AND RECOMMENDATIONS

Summary

This research was an empirical investigation into the effectiveness of the TEDS method when integrated into an iterative evaluation/treatment generation process for information artifact development. It was also a discovery and methodology-related study that produced insights about the effectiveness of the component parts of the methodological building blocks and how they can be combined. It also led to insights about the consistency of TEDS when scaling the TEDS sub-categories through clustering.

The study investigated the effectiveness and efficiency of an assessment approach to developing web-based information artifacts based on the TEDS assessment framework, iterative feature analysis and modification, and A/B testing. It was designed to provide insights into how the TEDS model could be used to generate feature improvement treatments based on TEDS comparative rankings, and the assessment comments and screenshots to generate feature modification treatments. In addition, it aimed to demonstrate how integrating TEDS with A/B testing could provide quantitative experimentation data to inform whether the proposed feature measurably improves the information artifact.

The study was a real-life example of how an organization whose core business is not IT used the TEDS framework and A/B testing in an iterative fashion. The method identified a prioritized list of feature improvement areas for the Sounders FC iOS mobile application in relation to 10 other professional sports iOS mobile applications. It provided a method to generate feature treatments from the TEDS ratings and the rater comments and screenshots. The method was then applied to the Sounders FC iOS mobile application to create feature improvement treatment specifications to address weaknesses of the application. The Sounders FC development team built two treatments from the specifications. The treatments were A/B tested using Yahoo Flurry, which is a basic data analytic platform that is available for free on the web. Both treatments were deployed to all users of the application based on the criteria determined by the Sounders development organization.

One challenge of doing a study embedded in a development organization over an extended period is it is difficult to avoid changes made to the information artifacts by the organization. In addition to the specific treatments that were A/B tested, the Sounders FC development team made at least 138 code changes (table 36) to the application that had an impact on the subsequent TEDS evaluations. Based on the development release notes, it appears that the Sounders made larger development investments in other scenarios during the project, which the improved TEDS ratings also reflected. However, this study showed that software updates that were tied to specific TEDS attributes in the Sounders FC iOS application made significant improvements in the Phase 2 TEDS assessments.

Since the Sounders FC development team made many changes to the mobile application besides the Player Information treatments that were A/B tested, the Phase 2 scenario ratings for Player Information do not provide as much of a "wow" factor in exceeding the improvement of other scenarios. However, this does not mean that the process was ineffective. Instead, it shows that the few changes recommended by the process had a positive impact on the scenarios, as evidenced by the subsequent TEDS ratings. And, as expected, the other changes that the Sounders FC development team deployed to address weaknesses they identified also had a positive impact on the scenarios, as evidenced by the subsequent TEDS ratings.

It was found that the treatment specification method proposed was effectively guided by the TEDS assessment data and the A/B testing. This shows that the methodological approach demonstrated the robustness and capacity to guide development in a targeted way, and to address weaknesses that are exposed through the TEDS evaluations. This makes this mixed method uniquely effective, even considering new literature and other frameworks.

The TEDS method provided ranked ratings of the artifacts for each scenario. This ranking was used to prioritize the screenshots and comments for the treatment improvement specifications. organizing the TEDS evaluation data in a way that elevated the best relevant information helps designers during the brainstorming process. Designs still needed to be wireframed based on the organized screenshots and comments.

One of the arguments for using A/B testing in a development organization is it provides a scientific way to listen to customers about features, rather than the opinions of those with decision-making power in an organization. The term *hippo* (Kohavi et al., 2007) refers to the highest paid person in an organization. Without behavioral user data, organizations often rely on the hippo or the most influential person in the room to decide the merits of building and deploying a new feature. In contrast, data-driven organizations use metrics from users (such as those included in a well-designed A/B experiment) to determine the merits of a new feature compared to an existing feature, or to compare changes to an existing feature with the current implementation of that feature. Similarly, a development organization can use metrics to prioritize feature areas or scenarios to invest in to be more competitive.

General Recommendations

Organizations that do not know how to evaluate their information artifacts, use A/B testing, or have a systematic way to prioritize and evaluate feature improvements can use this study as a template to iteratively assess their information artifacts and to optimize their development investments for design patterns that are identified in well-performing artifacts in the same competitive landscape. The TEDS method compels an organization to determine other artifacts that are similar in purpose and to define the scenarios and personas that are important for the IT artifact. Artifacts that are included in a study can come from competitors or from other successful artifacts that align, if the artifacts meet similar user needs and can be evaluated for similar user scenarios.

After the initial TEDS assessments are completed, the organization has a baseline of rankings for their artifact relative to the other artifacts in each scenario. The iterative process for A/B testing and IT assessment does not require that other artifacts are included in each iteration. Value exists in comparing how the artifact rates over time. For example, in this project, Phase 1 evaluations included 10 artifacts. The assessment data for this initial assessment could produce many treatment ideas across the eight scenarios. The Sounders FC development team could have continued A/B testing treatments based on this initial data for some time. In contrast, the Phase 4 TEDS assessment included only evaluations for the Sounders artifact. If they chose to, the Sounders could gain longitudinal insight by reevaluating the TEDS assessments of their iOS artifact in the future.

Based on insights from this study, it is recommended that organizations use the full TEDS method periodically, and then use the data generated to create feature treatments for A/B testing. Then, as the A/B testing is performed and some of the treatments are deployed, organizations can perform supplemental TEDS assessments on their artifact to gather information to see if user ratings improve. For example, the Sounders FC application development cycle corresponds to a yearly schedule because of the nature of the MLS soccer season. Each year, a full TEDS study could be performed that includes other artifacts from other professional teams. The output from the full TEDS project could be used to generate feature improvement specifications that would supplement the Sounders FC development team's development backlog. Throughout the season, the Sounders could continue to gather TEDS assessments of the evolving application to confirm that users rate the improvements that were recommended by the winning A/B experiments.

For IT companies that regularly use A/B testing, this study demonstrates how they can gain finer-grained insight into which features they should A/B test by iteratively applying an assessment framework like TEDS in combination with A/B testing. To produce examples of other implementations from which to leverage design patterns, the treatment generation process requires that other competitive information artifacts are included in the TEDS assessment phase.

Future Research

Clustering TEDS values was proven to be effective. But determining which criteria are most pertinent for evaluating an IT artifact in a specific context would be insightful. Sub-criteria could be ranked based on relevance for the artifact in a given context. In that way, the number of attributes could be increased or decreased, depending on the requirements of the researcher.

The methodology used for investigating attributes of a given artifact or context could be performed in several ways. For example, an instrument could be created to survey users about the importance of individual attributes. Or, analysis of existing TEDS assessment studies might provide insight based on analysis of the ratings and standard deviation of the individual attributes between artifacts, where the average rating between the artifacts is large.

The literature review in this dissertation compares the TEDS framework to several important assessment frameworks, yet the theory underlying some of these frameworks continues to refine and evolve in the literature. A refresh of the comparison between framework attributes, along with the TEDS analysis of how the TEDS clusters correspond, would be beneficial.

The emergence of applied evaluations using other theoretical frameworks in the literature makes it possible to apply learnings from the research to the TEDS method. The studies that developed an online instrument to survey user evaluations of information artifacts should be analyzed to improve the TEDS instrument used with users. For example, analysis of the wording used to operationalize criterion from the surveys of other frameworks can be compared to the wording currently used in TEDS when the criterion overlap frameworks.

Other aspects of online surveys to assess quality should be investigated. In the four data sets of user surveys completed in this dissertation, an interesting pattern emerged: User and expert ratings for the clusters which showed earlier in the survey, was lower than the ratings for clusters evaluated later. The question is whether this difference is due to the evaluation by the raters or the ordering of the clusters. Research should be done to see if randomizing the order of the cluster questions in the survey impacts the ratings by users.

Empirical analysis of the TEDS criteria using quantitative methods should be done. The initial dissertation proposal to use historical experiments from Microsoft was withdrawn because the metadata concerning the experiments was missing. However, an A/B experimentation platform could remedy this issue by providing a tagging system based on the TEDS framework. Other mechanisms in the A/B platform could support better experimental data gathering, such as ways to attach screenshots and code differences to the experiment for future analysis. Automation of the metadata gathering process—without investment from the experimenter—is preferred because it eliminates any extra work that might be required.

A subsequent study will compare user and expert ratings that use the TEDS method. More work is needed with the user survey instrument before there is confidence that the assessments are reliable. Understanding the difference between user ratings and expert ratings can shed light on steps to improve the gathering of user data.

Reference

Alsabawy, A. Y., Cater-Steel, A., & Soar, J. (2016). Determinants of perceived usefulness of e-learning systems. *Computers in Human Behavior, 64*, 843-858. doi:http://dx.doi.org/10.1016/j.chb.2016.07.065

Alves, J. M., Savaris, A., von Wangenheim, C. G., & von Wangenheim, A. (2016). *Software Quality Evaluation of the Laboratory Information System used in the Santa Catarina State Integrated Telemedicine and Telehealth System.* Paper presented at the Computer-Based Medical Systems (CBMS), 2016 IEEE 29th International Symposium

Assila, A., & Ezzedine, H. (2016). Standardized Usability Questionnaires: Features and Quality Focus. *Electronic Journal of Computer Science and Information Technology, 6*(1).

Ballou, D., Madnick, S., & Wang, R. (2003). Special Section: Assuring Information Quality. *Journal of Management Information Systems, 20*(3), 9-11.

Barnes, S., & Vidgen, R. (2000). WebQual: An Exploration of Web-site Quality. *School of Management, University of Bath*, 8.

Chaffey, D. (2016). Which are the most popular A/B testing tools?

Clifton, B. (2008). *Advanced Web metrics with Google Analytics*. Indianapolis, Ind.: Wiley Pub.

Coyne, I. T. (1997). Sampling in qualitative research. Purposeful and theoretical sampling; merging or clear boundaries? *Journal of Advanced Nursing, 26*(3), 623-630.

Crook, T., Frasca, B., Kohavi, R., & Longbotham, R. (2009). Seven pitfalls to avoid when running controlled experiments on the web. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1105-1113.

Davis, F. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly, 13*(3).

DeLone, W. H., & McLean, E. R. (1992). Information Systems Success: The Quest for the Dependent Variable. *Information Systems Research, 3*(1), 60-95.

DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: a ten-year update. *Journal of Management Information Systems, 19*(4), 9-30.

Eapen, B. R., & Chapman, B. (2015). Mobile Access to ClinicalConnect: A User Feedback Survey on Usability, Productivity, and Quality. *JMIR mHealth and uHealth, 3*(2), e35. doi:10.2196/mhealth.4011

Eisenberg, B., Quarto-von Tivadar, J., & Davis, L. T. (2008). *Always be testing the complete guide to Google website optimizer*. Indianapolis, Ind.: Wiley Technology Pub.

Facca, F. M., & Lanzi, P. L. (2005). Mining interesting knowledge from weblogs: a survey. *DATA AND KNOWLEDGE ENGINEERING, 53*(3), 225-241.

File, T., & Ryan, C. (2014). Computer and internet use in the United States: 2013. *American Community Survey Reports*.

Forsgren, N., Durcikova, A., Clay, P. F., & Wang, X. (2016). The integrated user satisfaction model: Assessing information quality and system quality as second-order constructs in system administration. *Communications of the Association for Information Systems, 38*, 803-839.

Fourie, I., & Bothma, T. (2007). Information seeking: an overview of web tracking and the criteria for tracking software. *ASLIB PROCEEDINGS, 59*(3), 264-284.

Goodhue, D. L., & Thompson, R. L. (1995). Task-Technology Fit and Individual Performance. *MIS Quarterly, 19*(2), 213-236. doi:10.2307/249689

Gould, J. D., & Lewis, C. (1985). Designing for usability: key principles and what designers think. *Commun. ACM, 28*(3), 300-311. doi:http://doi.acm.org/10.1145/3166.3170

Hamid, A. A., Razak, F. Z. A., Bakar, A. A., & Abdullah, W. S. W. (2016). The Effects of Perceived Usefulness and Perceived Ease of Use on Continuance Intention to Use E-Government. *Procedia Economics and Finance, 35*, 644-649. doi:http://dx.doi.org/10.1016/S2212-5671(16)00079-4

Hanington, B. M. (2010). Relevant and Rigorous: Human-Centered Research and Design Education. *Design Issues, 26*(3), 18-26.

Hartmann, J., De Angeli, A., & Sutcliffe, A. (2008). *Framing the user experience: information biases on website quality judgement*. Paper presented at the Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems, Florence, Italy.

Hartmann, J., Sutcliffe, A., & De Angeli, A. (2007). *Investigating attractiveness in web user interfaces*. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems, San Jose, California, USA.

Hartmann, J., Sutcliffe, A., & De Angeli, A. (2008). Towards a theory of user judgment of aesthetics and user interface quality. *ACM Trans. Comput.-Hum. Interact., 15*(4), 1-30. doi:http://doi.acm.org/10.1145/1460355.1460357

Hays, R. D., & Revicki, D. (2005). Reliability and validity (including responsiveness). *Assessing quality of life in clinical trials, 2*, 25-39.

Hsu, C.-C., & Sandford, B. A. (2007). The Delphi technique: making sense of consensus. *Practical assessment, research & evaluation, 12*(10), 1-8.

Hussain, A., Mkpojiogu, E. O., Yusof, M. M., Nifa, F. A. A., Nawi, M. N. M., & Hussain, A. (2016). *Perceived usefulness, perceived ease of use, and perceived enjoyment as drivers for the user acceptance of interactive mobile maps.* Paper presented at the AIP Conference Proceedings.

Ivory, M. Y., & Hearst, M. A. (2001). The state of the art in automating usability evaluation of user interfaces. *ACM Comput. Surv., 33*(4), 470-516. doi:http://doi.acm.org/10.1145/503112.503114

Jurisch, M., Krcmar, H., Scholl, H. J., Wang, K., Wang, Y., Woods, G., . . . Yao, Y. (2014, 6-9 Jan. 2014). *Digital and Social Media in Pro Sports: Analysis of the 2013 UEFA Top Four.* Paper presented at the 2014 47th Hawaii International Conference on System Sciences.

Kaplan, A. M. (2012). If you love something, let it go mobile: Mobile marketing and mobile social media 4x4. *Business Horizons, 55*(2), 129-139. doi:http://dx.doi.org/10.1016/j.bushor.2011.10.009

Kaushik, A. (2006). Experimentation and Testing: A Primer. Retrieved from https://www.kaushik.net/avinash/experimentation-and-testing-a-primer

Kaushik, A. (2007). *Web analytics : an hour a day*. Indianapolis, Ind.: Sybex.

Kaushik, A. (2010). *Web analytics 2.0 the art of online accountability & science of customer centricity*. Hoboken, N.J.: Wiley.

Klischewski, R., & Scholl, H. J. (2008). Information quality as capstone in negotiating e-government integration, interoperation and information sharing. *Electron. Gov. Electronic Government, 5*(2), 203-225.

Kohavi, R. (2012). *Online controlled experiments*. ACM, 2 Penn Plaza, Suite 701, New York, NY 10121-0701, USA.

Kohavi, R., Deng, A., Frasca, B., Walker, T., Xu, Y., & Pohlmann, N. (2013). *Online controlled experiments at large scale*. ACM, 2 Penn Plaza, Suite 701, New York, NY 10121-0701, USA.

Kohavi, R., Henne, R., & Sommerfield, D. (2007). *Practical guide to controlled experiments on the web: listen to your customers not to the hippo*. Paper presented at the Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, San Jose, California, USA.

Kohavi, R., & Longbotham, R. (2011). Unexpected results in online controlled experiments: building seemingly predictive models on random data. *SIGKDD Explor. Newsl. ACM SIGKDD Explorations Newsletter, 12*(2), 31-35.

Kohavi, R., Longbotham, R., Sommerfield, D., & Henne, R. (2009). Controlled experiments on the web: survey and practical guide. *Data Mining and Knowledge Discovery, 18*(1), 140-181.

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences, 111*(24), 8788-8790. doi:10.1073/pnas.1320040111

Lee, Y. W., Strong, D. M., Kahn, B. K., & Wang, R. Y. (2002). AIMQ: a methodology for information quality assessment. *Information & Management, 40*(2), 133-146.

Legner, C., Urbach, N., & Nolte, C. (2016). Mobile business application for service and maintenance processes: Using ex post evaluation by end-users as input for iterative design. *Information & Management, 53*(6), 817-831. doi:http://dx.doi.org/10.1016/j.im.2016.03.001

Lillrank, P. (2003). The quality of information. *International Journal of Quality & Reliability Management, 20*(6), 691-703.

McFarland, C. (2013). *Experiment! website conversion rate optimization with A/B and multivariate testing*. Berkeley, Calif.: New Riders.

McKinney, V., Yoon, K., & Zahedi, F. (2002). The measurement of web-customer satisfaction: An expectation and disconfirmation approach. *Information Systems Research, 13*(3), 296-315.

Morgan-Thomas, A. (2016). Rethinking technology in the SME context: Affordances, practices and ICTs. *International Small Business Journal, 34*(8), 1122-1136. doi:doi:10.1177/0266242615613839

Murphy, J., Hofacker, C. F., & Bennett, M. (2001). Website-generated Market-research Data: Tracing the Tracks Left Behind by Visitors. *Cornell Hotel & Restaurant Administration Quarterly, 42*(1).

Nabavi, M., & Jamali, H. R. (2015). Adding value to information systems: A new model. *Business Information Review, 32*(1), 53-59.

Nguyen, T. H., Newby, M., & Macaulay, M. J. (2015). Information Technology Adoption in Small Business: Confirmation of a Proposed Framework. *Journal of Small Business Management, 53*(1), 207-227.

Okoli, C., & Pawlowski, S. D. (2004). The Delphi method as a research tool: an example, design considerations and applications. *Information & Management, 42*(1), 15-29.

Peral, J., Maté, A., & Marco, M. (2016). Application of Data Mining techniques to identify relevant Key Performance Indicators. *Computer Standards & Interfaces*. doi:https://doi.org/10.1016/j.csi.2016.11.006

Pullen, P. C. (2010). Understanding Standard Scores. Retrieved from http://faculty.virginia.edu/PullenLab/WJIIIDRBModule/WJIIIDRBModule7.html

Rai, A., Lang, S. S., & Welker, R. B. (2002). Assessing the validity of IS success models: An empirical test and theoretical analysis. *Information Systems Research, 13*(1), 50-69.

Rogers, A. D. (2016). *Examining Small Business Adoption of Computerized Accounting Systems Using the Technology Acceptance Model.* (PhD), Walden University, ProQuest Dissertations Publishing. (10001635)

Rubin, J. (Ed.) (1994). *Handbook of usability testing: How to plan, design, and conduct effective tests*. New York: Wiley.

Sackman, H. (1974). *Delphi assessment: Expert opinion, forecasting, and group process* (S. M. Rand, Calif. Ed.): DTIC Document.

Saghaeiannejad-Isfahani, S., Saeedbakhsh, S., Jahanbakhsh, M., & Habibi, M. (2015). Analysis of the quality of hospital information systems in Isfahan teaching hospitals based on the DeLone and McLean model. *Journal of Education and Health Promotion, 4,* 5. doi:10.4103/2277-9531.151883

Samardžija, A. Ć. (2016). *Measuring the Success of the Interactive Mobile Information Systems at the Individual Level of Use* (PhD), University of Zagreb, University of Zagreb.

Schmidt, R. C. (1997). Managing Delphi Surveys Using Nonparametric Statistical Techniques. *Decision Sciences, 28*(3).

Scholl, H. J. (2013). Evaluating sports websites from an information management perspective. *Routledge Handbook of Sport Communication*, 289.

Scholl, H. J., & Carlson, T. S. (2012). Professional sports teams on the Web: A comparative study employing the information management perspective. *European Sport Management Quarterly, 12*(2), 137-160.

Scholl, H. J., Eisenberg, M. B., Dirks, L., & Carlson, T. S. (2011a). The TEDS framework for assessing information systems from a human actors' perspective: Extending and repurposing Taylor's Value-Added Model. *Journal of the American Society for Information Science and Technology, 62*(4), 789-804. doi:10.1002/asi.21500

Scholl, H. J., Eisenberg, M. B., Dirks, L., & Carlson, T. S. (2011b). The TEDS framework for assessing information systems from a human actors' perspective: Extending and repurposing Taylor's Value-Added Model. *Journal of the American Society for Information Science & Technology, 62*(4).

Scholl, H. J., Wang, K., Wang, Y., Woods, G., Xu, D., Yao, Y., . . . Krcmar, H. (2014). Top soccer teams in cyberspace: Online channels for services, communications, research, and sales. *Journal of Marketing Analytics, 2*(2).

Scholl, M., Ehrlich, P., Wiesner-Steiner, A., & Edich, D. (2014). *The project TEDS@ wildau: TEDS Framework Integration into the Moodle platform for user-specific quality assurance of learning scenarios.* Paper presented at the System Sciences (HICSS), 2014 47th Hawaii International Conference on.

Sharma, G., & Lijuan, W. (2015). The effects of online service quality of e-commerce Websites on user satisfaction. *The Electronic Library, 33*(3), 468-485. doi:10.1108/el-10-2013-0193

Shchiglik, C., Barnes, S. J., & Scornavacca, E. (2016). The Development of an Instrument to Measure Mobile Game Quality. *Journal of Computer Information Systems, 56*(2), 97-105. doi:10.1080/08874417.2016.1117368

Skulmoski, G. J., Hartman, F. T., & Krahn, J. (2007). The Delphi Method for Graduate Research. *Journal of Information Technology Education, 6*, 1-21.

Tang, D., Agarwal, A., O'Brien, D., & Meyer, M. (2010). *Overlapping experiment infrastructure*. ACM, 2 Penn Plaza, Suite 701, New York, NY 10121-0701, USA.

Taylor, R. S. (1982). Value-Added Processes in the Information Life Cycle. *Journal of the American Society for Information Science, 33*(5), 341-346. doi:10.1002/asi.4630330517

Taylor, R. S. (1986). *Value-added processes in information systems*. Norwood, N.J. :: Ablex Pub. Corp.

Wand, Y., & Wang, R. Y. (1996). Anchoring data quality dimensions in ontological foundations. *Commun. ACM, 39*(11), 86-95. doi:http://doi.acm.org/10.1145/240455.240479

Wixom, B. H., & Todd, P. A. (2005). A theoretical integration of user satisfaction and technology acceptance. *Information Systems Research, 16*(1), 85-102.

Yang, Z., Cai, S., Zhou, Z., & Zhou, N. (2005). Development and validation of an instrument to measure user perceived service quality of information presenting Web portals. *Information & management., 42*(4), 575.

APPENDICES

Appendix A: TEDS Variables

Term	Definition
Aggregate Scenario Team Total (ASTT)	The sum of all value ratings of all user scenarios for a given artifact. When all TEDS criteria are included, there are 40 criteria. The ASTT can also be calculated by adding all the Criteria Totals (CTs) for a given IT artifact.
Criteria Total (CT)	The sum of all values for specific user criteria. The number of values varies for each of the 6 user criteria in a TEDS study. For example, Ease of Use has 7 values (Browsing/browsability/searchability, Formatting/presentation, Mediation, Orientation, Order/consistency, Accessibility, Simplicity). The CT is the sum of the Likert ratings for the seven values. Similarly, Performance has four values (Cost savings, Time savings, Security, Safety), so its CT is the sum of the Likert ratings for the four values.
Scenario Criteria Total (SCT)	The sum of each artifact's CTs for a given user scenario. For example, for example, the SCT is derived by adding the CTs for Artifact1, Artifact2, and the other artifacts included in the study.
Aggregate of Criteria for given Scenario Total (ACST)	The sum of SCTs for all the user criteria for one user scenario. For example, Ease of Use SCT + Noise Reduction SCT + Quality SCT + Adaptability SCT + <others>.
Aggregate of Scenarios for given Criteria Total (ASCT)	The sum of a user criterion's SCTs for all user scenarios. For example, for Ease of Use, it is (Ease of Use SCT for Scenario1) + (Ease of Use SCT for Scenario2) + (Ease of Use SCT for Scenario3) + (Ease of Use SCT for <others>).
Weighted Average Criteria Total (WACT)	The weighted average of a user criterion which is found by dividing the CT of a user criterion by its number of values. For example, for Scenario1, if the Quality has Likert ratings of: contextuality = 2, flexibility = 3, simplicity = 3, privacy = 2, the WACT for Quality is (2 + 3 +3 +2) / 4, or 2.5.

Weighted Average of Scenario Criteria Total (WASCT)	The sum of all team WACTs for a given user scenario. For example, for Scenario1, adding the WACTs for Artifact1, Artifact 2, and the other artifacts gives the WASCT.
Aggregate of Weighted Average Criteria for given Scenario Total (AWACST)	The sum of WASCTs for all the user criteria for a given user scenario, such as (Ease of Use WASCT) + (Noise Reduction WASCT) + <Quality WASCT> + <others>.
Aggregate of Scenarios for a given Weighted Average Criteria Total (ASWACT)	The sum of a given user criterion's WASCTs for all user scenarios. For example, for Ease of Use, it is (Ease of Use WASCT for Scenario1) + (Ease of Use WASCT for Scenario2) + (Ease of Use WASCT for Scenario3) + (WASCT for <other scenarios>).
Team Criteria Total (TCT)	The sum of a given team's CTs across all user scenarios. For example, Artifact1 has a CT for each of scenarios. The Artifact1 TCT for Ease of Use is the sum of the Artifact1 Ease of Use CTs for the eight scenarios: (Artifact1 Ease of Use for Scenario1) + (Artifact1 Ease of Use for Scenario2) + (Artifact1 of Use for Scenario3) + (Artifact1 Ease of User for <remaining other scenarios>)
Aggregate of Criteria for given Team Total (ACTT)	The sum of a given team's TCTs for all user criteria. For example, for Artifact1, the ACTT is: (Ease of Use TCT for Artifact1) + (Noise Reduction TCT for Artifact1) + (Quality TCT for Artifact1) + <others>.
Aggregate of Teams for given Criteria Total (ATCT)	The sum of the user criteria TCTs for all teams. For example, for the Ease of Use criteria, the ATCT is: (Ease of Use TCT for Artifact1) + (Ease of Use TCT for Artifact2) + (Ease of Use TCT for Artifact3) + <others>.
Weighted Average of Team Criteria Total (WATCT)	The sum of all scenario WACTs for all the user criteria for a given team. For example, for Artifact1, the WATCT is: (Scenario1 WACT) + (Scenario2 WACT) + (Scenario3 WACT) + <others>).
Aggregate of Weighted Average Criteria for given Team Total (AWACTT)	The sum of WATCTs for all the user criteria for a given team: (Ease of Use WATCT + Noise Reduction WATCT + <others>).

Aggregate of Teams for a given Weighted Average Criteria Total (ATWACT)	The sum of a given user criteria's WATCTs for all user scenarios. For example, for *Ease of Use*, it is: (*Ease of Use* WATCT for Scenario1) + (*Ease of Use* WATCT for Scenario2) + (*Ease of Use* WATCT for Scenario3) + (*Ease of Use* WATCT for <*others*>).

Notes

ASTT = ACTT

The average Likert rating is easily calculated from the ASTT. It is ASTT/40 (where 40 is the number of criteria in the TEDS framework).

When using the average Likert rating to rank scenarios, teams, or user criteria, it will correlate perfectly to the non-weighted totals (such as ASTT) of the summed data being evaluated. The same is not always true for the ranking of weight-averaged summed data.

Appendix B: Comparison Summary of Assessment Frameworks (page 1 of 4)

Reference	Framework	Description	Domain	Year
Taylor	**Value Added Model**	Information Systems add the following values to information for users: • Browsing • Formatting • Mediation • Orienting • Ordering • Physical Accessibility • Access I (Item Identification) • Access II (Subject Description) • Access III (Subject Summary) • Linkage • Precision • Selectivity • Accuracy • Comprehensiveness • Currency • Reliability • Validity • Closeness to Problem • Flexibility • Simplicity • Stimulatory • Time Savings • Cost Savings	Evaluation and Assessment	1980

Appendix B: Comparison Summary of Assessment Frameworks (page 2 of 4)

Reference	Framework	Description	Domain	Year
Davis (1989) Chan and Teo (2007) Lederer, et al. (2000) Chuan-Chuan Lin and Lu (2000)	**Technology Acceptance Model (TAM)**	PU and PEOU Davis shows that U and PEOU predict user acceptance of behavior	User Acceptance	1989
DeLone and McLean (1992) Seddon (1997)	**DeLone and McLean Information Success Model: (D&M Success)**	• System Quality • Information Quality • IS Use • User Satisfaction • Individual Impact • Organizational Impact Success measures interrelated	User Acceptance	1992
Wand and Wang	**Data Quality Dimensions in Ontological foundations**	• Accuracy and Precision • Reliability • Timeliness & Currency • Completeness • Consistency	Evaluation and Assessment	1996
Barnes and Vidgen	**WebQual**	24 attributes in 4 categories • Ease of use • Experience • Information • Communication and Integration	Evaluation and Assessment	2000
Lee, Strong, Kahn and Wang	**Aim Quality (AIMQ)**	• PSP/IQ 2x2 framework of what IQ means to users • IQA questionnaire for measuring IQ • IQ Gap Analysis techniques	Evaluation and Assessment	2002

Appendix B: Comparison Summary of Assessment Frameworks (page 3 of 4)

Reference	Framework	Description	Domain	Year
McKinney, Yoon, Zahedi	**IS/SQ Disconfirmation Model**	Information Quality (SQ) • Relevance • Timeliness • Reliability • Scope • Perceived Usefulness System Quality (SQ) • Access • Usability • Attractive • Navigation • Interactivity uses "disconfirmation" to measure user satisfaction (expectation versus perceived performance)	Evaluation and Assessment	2002
Yang, Cai, Zhou, Zhou	**Five Dimension Service Quality Instrument**	• Usability • Usefulness of content • Adequacy of information • Accessibility • Interaction	Evaluation and Assessment	2005
Hartmann, De Angeli, Sutcliffe	**Adaptive Decision Making (ADM)**	• Usability • Content • Aesthetics • Reputation/identity • Customizability	Evaluation and Assessment	2008

Appendix B: Comparison Summary of Assessment Frameworks (page 4 of 4)

Reference	Framework	Description	Domain	Year
Scholl, H, J, et al	**TEDS (Taylor revised)**	Ease of Use • Browsing/searchability • Formatting • Mediation • Orientation • Order/consistency • Accessibility • Simplicity Noise Reduction • Item identification • Subject description • Subject summary • Linkage/referral • Precision • Selectivity • Order • Novelty Quality • Accuracy • Comprehensiveness • Currency • Reliability • Validity • Authority Adaptability • Contextuality • Flexibility • Simplicity • Transaction • Trust • Feedback • Community • Individualization • Localization • Privacy Performance	Evaluation and Assessment	2011

		• Cost Savings • Time Savings • Security • Safety Affection • Aesthetics • Entertainment • Engagement • Stimulation • Satisfaction		

Appendix C: TEDS Framework

Table 6: Definitions of User Criteria Values

VALUES	Description/Definition
Accuracy	The "value-added by system processes that assures error-free transfer of data and information as it flows through the IS/IT artifact and is eventually displayed to a human actor" (Taylor, 1986, p. 70). Exactness of a piece of information and its conformity to the original source as well as the exact and unaltered transfer and presentation of a piece of information to a human actor through an IS/IT artifact.
Comprehensiveness	The "value-added by the completeness of coverage of a particular subject or of a particular form of information" (Taylor, 1986, p. 70). Complete and broad coverage of a particular subject, in which a human actor is interested, provided by a source/piece of information or an IS/IT artifact.
Currency	The "value-added (a) by the recency of the data acquired by the system; and (b) by the capability of the IS/IT artifact to reflect current modes of thinkingin its structure, organization, and access vocabularies" (Taylor, 1986, p. 70). The recency of a source/piece of information sought and the capability of an IS/IT artifact to reflect and represent current modes of understanding of the sought subject matter to a human actor.
Reliability	The "value-added by the trust an IS/IT artifact inspires in human actors by its consistency of quality performance over time" (Taylor, 1986, p. 70). The suitability and dependability of a source/piece of information or of an IS/IT artifact that a human actor consistently experiences.
Validity	The "value-added when the IS/IT artifact provides signals about the degree to which data or information presented to human actors can be judged as sound" (Taylor, 1986, p. 70). The quality of a source/piece of information or of an IS/IT artifact to be assessed as sound, justifiable, well-grounded, and logically correct by a human actor.
Authority	The extent of credibility and reputation human actors attribute to the human, technical, or institutional sources/pieces of information or to an IS/IT artifact.

Table 7: Ease of Use Criteria

Values	Description/Definition
Browsing/browsability/searchability	Human actor-expected capability of an IS/IT artifact to "scan an information neighborhood with the probability that the human actor will serendipitously find information of value" (Taylor, 1986, p. 70)
Formatting/presentation	The "physical presentation and arrangement of data/information in ways that allow more efficient scanning and hence extraction of items of interest from the store" (Taylor, 1986, p. 70). Presenting and arranging information to facilitate scanning and selecting (Choo, 2002) on part of a human actor.
Mediation	The "means used to assist human actors in getting answers from the IS/IT artifact" (Taylor, 1986, p. 70). The act or process of intervening between the IS/IT artifact and the human actor to promote reconciliation or understanding, that is, the intermediation between the functions and characteristics of the IS/IT artifact and the human actor.
Orientation	The "means used to help human actors understand and to gain experience with the IS/IT artifact and its complexities" (Taylor, 1986, p. 70). The ease with which a human actor can orient himself/herself regarding the intended or emerging utilization of the IS/IT artifact at hand; that is, the IS/IT artifact gives clear clues to where the human actor can find information or functions such as intuitive navigation, breadcrumbs, and other visual and audio clues.
Order/consistency	The "value-added by initially dividing or organizing a body of subject matter by some form of gross ordering, such as alphabetization, or large groupings" (Taylor, 1986, p. 70). Orderly and systematic arrangement of IS/IT artifact components and elements such that a human actor can easily identify information and functionality provided by the IS/IT artifact. Also, the fashion through which the IS/IT artifact transforms and presents itself based on human action.
Accessibility	The "processes of making access to information stores easier in a physical sense" (Taylor, 1986, p. 70). Ease of physical access to an IS/IT artifact and to the information and functionality that the IS/IT artifact can provide to human actors (including to impaired human actors). Also, how many, or how few, steps does the human actor need (for example, in terms of clicks) to reach the desired information or functionality.
Simplicity	Lack of complexity, complication, or difficulty when a human actor interacts with and operates an IS/IT artifact.

Table 8. Performance Criteria

VALUES	Description/Definition
Cost savings	The "value achieved by conscious IS/IT artifact design and operating decisions that save dollars for the human actor" (Taylor, 1986, p. 70). The extent to which a human actor can cut real cost when utilizing a source/piece of information or an IS/IT artifact.
Time savings	The "perceived value of an IS/IT artifact based on the speed of its response time" (Taylor, 1986, p. 70). The extent to which a human actor can save time when utilizing a source/piece of information or an IS/IT artifact.
Security	The extent to which a source/piece of information or an IS/IT artifact provides safeguards against fraud and intrusion such that a human actor can feel secure, protected, and free of anxiety when utilizing a source/piece of information or an IS/IT artifact.
Safety	The extent to which a source/piece of information or an IS/IT artifact provides safeguards against the risk of hurt, injury, loss, or danger such that a human actor can feel safe when utilizing a source/piece of information or an IS/IT artifact.

Table 9. Noise Reduction Criteria

VALUES	Description/Definition
Item identification	The "value achieved by the identification of any information chunk or discrete piece of data by systematic physical description and location information" (Taylor, 1986, p. 69). A humanactor- oriented comprehensive description of the physical and functional characteristics as well as the location of a unit of information or of an IS/IT artifact.
Subject description/ classification/ controlled vocabulary	The "provision of a subject description through access points such as index terms, descriptors, and names" (Taylor, 1986, p. 70). Human-actor-oriented interpretive layer, which helps describe, arrange into groups, and classify information or physical and functional characteristics of an IS/IT artifact; that also includes that a human actor can easily identify a source/piece of information or an IS/IT artifact characteristic by its classification.
Subject summary/ summarization	The "result of processes which reduce or compress large amounts of information into compact items, such as executive summaries, abstracts, terse conclusions, chemical structure diagrams, mathematical formulae, graphs, or charts" (Taylor, 1986, p. 70). A brief summary or abstract of information or of the physical and functional characteristics of an IS/IT artifact.
Linkage/referral	The "value-added by providing pointers and links to items, sources, and systems external to the IS/IT artifact in use, thus expanding the human actor's information options" (Taylor, 1986, p. 70).
Precision/(Relevant Retrieved) over (Retrieved)	The "capability of an IS/IT artifact to aid human actors in finding exactly what they want, by providing signals on such attributes as language, data aggregation, sophistication level, or by ranking output" (Taylor, 1986, p. 70). The capacity of a source/piece of information or the physical or functional characteristics of an IS/IT artifact to provide a human actor exactly with what she or he expects and needs.
Selectivity	The "value-added when choices are made at the input point of the IS/IT artifact, choices based on the appropriateness and merit of information chunks to the client population served" (Taylor, 1986, p. 70). The number and the nature of choices a human actor encounters and chooses from when she or he uses a source/piece of information or an IS/IT artifact.
Order	Parsimony, noise-reduced structure, and absence of distraction that a human actor encounters when she or he uses a source/piece of information or an IS/IT artifact.
Novelty	The extent of originality and newness relative to a human actor's needs that she or he encounters when using a source/piece of information or an IS/IT artifact, that is, the balance of maintaining currency and overloading the human actor.

Table 10. Adaptability Criteria

VALUES	Description/Definition
Contextuality/closeness to problem	The "value-added by the activities of the system, usually through human intervention, to meet the specific needs of a person in a particular environment with a particular problem; this implies knowledge of that person's style, bias, idiosyncrasies, and sophistication, as well as the politics and constraints of the context" (Taylor, 1986, p. 70). The extent to which a source/piece of information or an IS/IT artifact matches a human actor's specific informational or transactional needs, also with respect to that human actor's specific location when stationary or on the move.
Flexibility	The "capability of an IS/IT artifact to provide a variety of ways and approaches of working dynamically with the data/information in a file" (Taylor, 1986, p. 70). The dynamic adjustment of information or an IS/IT artifact to a human actor's changing informational or transactional needs.
Simplicity	The "value achieved by presenting the most clear and lucid (explanation, data, hypothesis, or method) among several within quality and validity limits" (Taylor, 1986, p. 70). The lack of complication or difficulty with which information or an IS/IT artifact adjust to a human actor's changing informational and transactional needs.
Transaction	The capacity of an IS/IT artifact to immediately and on a per-demand basis cope with a human-actor's transactional need without referral or deferral (a transaction might include the purchasing of goods and services, electronic or nonelectronic).
Trust	An individual humanactor's willingness to consider a source/piece of information or an IS/IT artifact trustworthy and to act accordingly based on accumulated experience or other clues such as certificates, ratings, or reviews.
Feedback	The capacity of a source/piece of information or an IS/IT artifact to entertain, receive feedback from and display feedback to human actors who utilize and assess that source/piece of information or that IS/IT artifact.
Community/social networking	The capacity of a source/piece of information or an IS/IT artifact to help human actors form a community or social network, electronically or nonelectronically, around a set of shared interests.
Individualization	The capacity of a source/piece of information or an IS/IT artifact to adjust to an individual human-actor's specific needs. Two forms of individualization are distinguished: (a) Static or basic individualization, for example, based on a human actor's preset preferences and selections including push-updates of information; (b) advanced or dynamic individualization where specific human-actors' needs are addressed as those that emerge through utilization and relative to the changing contexts over time (including the inference of changes in human-actors' preferences, ambiences, or search patterns as well as suggesting the selection of information or the potential utilization of an IS/IT artifact).
Localization	The extent to which a source/piece of information or an IS/IT artifact is sensitive to or reflective of differences in physical measures and metrics, time zones, languages, cultural, and other differences in real time relative to a human-actor's specific needs.
Privacy	A human actor's right "to be left alone" (Warren & Brandeis, 1890) and to be or remain apart from company, observation, tracking, and recording of activities and selections when utilizing a source/ piece of information or an IS/IT artifact.

Table 11. Affection Criteria

VALUES	Description/Definition
Aesthetics	The extent to which a human actor appreciates the appearance and perceived beauty of presentation when using a source/piece of information or an IS/IT artifact.
Entertainment	The extent to which a human actor appreciates the perceived amusement and diversion of presentation and interaction when using a source/piece of information or an IS/IT artifact.
Engagement	The extent to which a human actor appreciates the attractiveness of presentation and the appeal of interaction when using a source/piece of information or an IS/IT artifact.
Stimulation	The extent to which a human actor feels stimulated to act or grow to greater activity when using a source/piece of information or an IS/IT artifact.
Satisfaction/rewarding/incenting	The extent to which a human actor feels personally satisfied, incented, and rewarded when using a source/piece of information or an IS/IT artifact.

Appendix D: TEDS Clustered Criteria Definitions

Cluster	**Definition**
Navigation and Findability (Ease of Use)	How easy is it to find the information you are looking for? Can you easily browse or search? Does the app provide filters to narrow down your search? This ranking refers to the information artifacts' effectiveness evaluated on the aspects of: • **Browsing/browsability/searchability** - Human actor-expected capability of an IS/IT artifact to scan an information neighborhood with the probability that the human actor will serendipitously find information of value. (Taylor, 1986, p. 70) • **Orientation** - The means used to assist human actors in getting answers from the IS/IT artifact. (Taylor, 1986, p. 70). The act or process of intervening between the IS/IT artifact and the human actor to promote reconciliation or understanding, that is, the intermediation between the functions and characteristics of the IS/IT artifact and the human actor. • **Mediation** - The means used to help human actors understand and to gain experience with the IS/IT artifact and its complexities. (Taylor, 1986, p. 70). The ease with which a human actor can orient himself/herself regarding the intended or emerging utilization of the IS/IT artifact at hand; that is, the IS/IT artifact gives clear clues to where the human actor can find information or functions such as intuitive navigation, breadcrumbs, and other visual and audio clues. • **Simplicity** - Lack of complexity, complication, or difficulty when a human actor interacts with and operates an IS/IT artifact.

Structure (Ease of Use)	How easy is it for you to understand how the information presented is ordered; and, how many interactions does it take to get the information you seek? This ranking refers to the information artifacts' effectiveness evaluated on the aspects of: • **Formatting/presentation** - The physical presentation and arrangement of data/information in ways that allow more efficient scanning and hence extraction of items of interest from the store. (Taylor, 1986, p. 70). Presenting and arranging information to facilitate scanning and selecting (Choo, 2002) on part of a human actor. • **Order/consistency** - The value-added by initially dividing or organizing a body of subject matter by some form of gross ordering, such as alphabetization, or large groupings. (Taylor, 1986, p. 70). Orderly and systematic arrangement of IS/IT artifact components and elements such that a human actor can easily identify information and functionality provided by the IS/IT artifact. Also, the fashion through which the IS/IT artifact transforms and presents itself based on human action. • **Accessibility** - The processes of making access to information stores easier in a physical sense. (Taylor, 1986, p. 70). Ease of physical access to an IS/IT artifact and to the information and functionality that the IS/IT artifact can provide to human actors (including to impaired human actors). Also, how many, or how few, steps does the human actor need (for example, in terms of clicks) to reach the desired information or functionality.

Identity (Noise Reduction)	How easy is it to identify items, based on their descriptions or subject categories? This ranking refers to the information artifacts' effectiveness evaluated on the aspects of: • **Item Identification** - The value achieved by the identification of any information chunk or discrete piece of data by systematic physical description and location information (Taylor, 1986, p. 69). A human actor- oriented comprehensive description of the physical and functional characteristics as well as the location of a unit of information or of an IS/IT artifact. • **Subject description/classification/controlled vocabulary** - The provision of a subject description through access points such as index terms, descriptors, and names (Taylor, 1986, p. 70). Human-actor-oriented interpretive layer, which helps describe, arrange into groups, and classify information or physical and functional characteristics of an IS/IT artifact; that also includes that a human actor can easily identify a source/piece of information or an IS/IT artifact characteristic by its classification. • **Subject summary/summarization** - The result of processes which reduce or compress large amounts of information into compact items, such as executive summaries, abstracts, terse conclusions, chemical structure diagrams, mathematical formulae, graphs, or charts (Taylor, 1986, p. 70). A brief summary or abstract of information or of the physical and functional characteristics of an IS/IT artifact. • **Precision/(Relevant Retrieved) over (Retrieved)** - The capability of an IS/IT artifact to aid human actors in finding exactly what they want, by providing signals on such attributes as language, data aggregation, sophistication level, or by ranking output (Taylor, 1986, p. 70). The capacity of a source/piece of information or the physical or functional characteristics of an IS/IT artifact to provide a human actor exactly with what she or he expects and needs. • **Selectivity** - The value-added when choices are made at the input point of the IS/IT artifact, choices based on the appropriateness and merit of information chunks to the client population served (Taylor, 1986, p. 70). The number and the nature of choices a human actor encounters and chooses from

	when she or he uses a source/piece of information or an IS/IT artifact.
<u>**Parsimony** (Noise Reduction)</u>	How well do artifacts provide links to other resources? Is the means to link to those resources novel? This ranking refers to the information artifacts' effectiveness evaluated on the aspects of: • **Linkage/referral -** The value-added when choices are made at the input point of the IS/IT artifact, choices based on the appropriateness and merit of information chunks to the client population served (Taylor, 1986, p. 70). The number and the nature of choices a human actor encounters and chooses from when she or he uses a source/piece of information or an IS/IT artifact. • **Order -** Parsimony, noise-reduced structure, and absence of distraction that a human actor encounters when she or he uses a source/piece of information or an IS/IT artifact. • **Novelty -** The extent of originality and newness relative to a human actors needs that she or he encounters when using a source/piece of information or an IS/IT artifact, that is, the balance of maintaining currency and overloading the human actor.

Completeness (Quality)	How accurate, comprehensive, and current do you find the information offered? This ranking refers to the information artifacts' effectiveness evaluated on the aspects of: • **Accuracy** - The value-added by system processes that assures error-free transfer of data and information as it flows through the IS/IT artifact and is eventually displayed to a human actor (Taylor, 1986, p. 70). Exactness of a piece of information and its conformity to the original source as well as the exact and unaltered transfer and presentation of a piece of information to a human actor through an IS/IT artifact. • **Comprehensiveness** - The value-added by the completeness of coverage of a particular subject or of a particular form of information (Taylor, 1986, p. 70). Complete and broad coverage of a particular subject, in which a human actor is interested, provided by a source/piece of information or an IS/IT artifact. • **Currency** - The value-added (a) by the recency of the data acquired by the system; and (b) by the capability of the IS/IT artifact to reflect current modes of thinking in its structure, organization, and access vocabularies (Taylor, 1986, p. 70). The recency of a source/piece of information sought and the capability of an IS/IT artifact to reflect and represent current modes of understanding of the sought subject matter to a human actor.
Trustworthiness (Quality)	How reliable, valid, and authoritative do you find the information provided? This ranking refers to the information artifacts' effectiveness evaluated on the aspects of: • **Reliability** - The value-added by the trust an IS/IT artifact inspires in human actors by its consistency of quality performance over time (Taylor, 1986, p. 70). The suitability and dependability of a source/piece of information or of an IS/IT artifact that a human actor consistently experiences. • **Validity** - The value-added when the IS/IT artifact provides signals about the degree to which data or information presented to human actors can be judged as sound (Taylor, 1986, p. 70). The quality of a source/piece of information or of an IS/IT artifact to be assessed as sound, justifiable, well-grounded, and logically correct by a human actor. • **Authority** - The extent of credibility and reputation human

	actors attribute to the human, technical, or institutional sources/pieces of information or to an IS/IT artifact.
Interaction (Adaptability)	How well does the system provide you with feedback; and, how easily are you able to consume and share articles of interest to your local network? This ranking refers to the information artifacts' effectiveness evaluated on the aspects of: • **Contextuality/closeness to problem** - The value-added by the activities of the system, usually through human intervention, to meet the specific needs of a person in a particular environment with a particular problem; this implies knowledge of that person's style, bias, idiosyncrasies, and sophistication, as well as the politics and constraints of the context (Taylor, 1986, p. 70). The extent to which a source/piece of information or an IS/IT artifact matches a human actors specific informational or transactional needs, also with respect to that human actors specific location when stationary or on the move. • **Transaction** - The capacity of an IS/IT artifact to immediately and on a per-demand basis cope with a human-actors transactional need without referral or deferral (a transaction might include the purchasing of goods and services, electronic or nonelectronic). • **Feedback** - The capacity of a source/piece of information or an IS/IT artifact to entertain, receive feedback from and display feedback to human actors who utilize and assess that source/piece of information or that IS/IT artifact. • **Community/social networking** - The capacity of a source/piece of information or an IS/IT artifact to help human actors form a community or social network, electronically or nonelectronically, around a set of shared interests.

Customization (Adaptability)	How well or easily does the app allow you to focus your interactions on local communities of shared interest, whether by using local languages or offering you a flexible means to narrow down content based on relevance to your location? This ranking refers to the information artifacts' effectiveness evaluated on the aspects of: • **Flexibility** - The capability of an IS/IT artifact to provide a variety of ways and approaches of working dynamically with the data/information in a file (Taylor, 1986, p. 70). The dynamic adjustment of information or an IS/IT artifact to a human actor changing informational or transactional needs. • **Simplicity** - The value achieved by presenting the most clear and lucid (explanation, data, hypothesis, or method) among several within quality and validity limits (Taylor, 1986, p. 70). The lack of complication or difficulty with which information or an IS/IT artifact adjust to a human actors changing informational and transactional needs. • **Trust** - An individual human actors willingness to consider a source/piece of information or an IS/IT artifact trustworthy and to act accordingly based on accumulated experience or other clues such as certificates, ratings, or reviews. • **Individualization** - The capacity of a source/piece of information or an IS/IT artifact to adjust to an individual human-actors specific needs. Two forms of individualization are distinguished: (a) Static or basic individualization, for example, based on a human actors preset preferences and selections including push-updates of information; (b) advanced or dynamic individualization where specific human-actors needs are addressed as those that emerge through utilization and relative to the changing contexts over time (including the inference of changes in human-actors preferences, ambiences, or search patterns as well as suggesting the selection of information or the potential utilization of an IS/IT artifact). • **Localization** - The extent to which a source/piece of information or an IS/IT artifact is sensitive to or reflective of differences in physical measures and metrics, time zones, languages, cultural, and other differences in real time relative to a human-actors specific needs. • **Privacy** - A human actors right to be left alone (Warren &

	Brandeis, 1890) and to be or remain apart from company, observation, tracking, and recording of activities and selections when utilizing a source/ piece of information or an IS/IT artifact.
Savings (Performance)	How well is the app structured to save you time in finding what you're interested in, in this scenario? This ranking refers to the information artifacts' effectiveness evaluated on the aspects of: • **Time Savings** - The value achieved by conscious IS/IT artifact design and operating decisions that save dollars for the human actor (Taylor, 1986, p. 70). The extent to which a human actor can cut real cost when utilizing a source/piece of information or an IS/IT artifact. • **Cost Savings** - The perceived value of an IS/IT artifact based on the speed of its response time (Taylor, 1986, p. 70). The extent to which a human actor can save time when utilizing a source/piece of information or an IS/IT artifact.
Confidence (Performance)	How comfortable do you feel using this app? This ranking refers to the information artifacts' effectiveness evaluated on the aspects of: • **Security** - The extent to which a source/piece of information or an IS/IT artifact provides safeguards against fraud and intrusion such that a human actor can feel secure, protected, and free of anxiety when utilizing a source/piece of information or an IS/IT artifact. • **Safety** - The extent to which a source/piece of information or

	an IS/IT artifact provides safeguards against the risk of hurt, injury, loss, or danger such that a human actor can feel safe when utilizing a source/piece of information or an IS/IT artifact.
Attractiveness (Affection)	This ranking refers to the information artifacts' effectiveness evaluated on the aspects of: • **Aesthetics** - The extent to which a human actor appreciates the appearance and perceived beauty of presentation when using a source/piece of information or an IS/IT artifact. • **Satisfaction/rewarding/incenting** - The extent to which a human actor feels personally satisfied, incented, and rewarded when using a source/piece of information or an IS/IT artifact.
Enjoyment (Affection)	How engaging and stimulating did you find the app, in this scenario? This ranking refers to the information artifacts' effectiveness evaluated on the aspects of: • **Entertainment** - The extent to which a human actor appreciates the perceived amusement and diversion of presentation and interaction when using a source/piece of information or an IS/IT artifact. • **Engagement** - The extent to which a human actor appreciates the attractiveness of presentation and the appeal of interaction when using a source/piece of information or an IS/IT artifact. • **Stimulation** - The extent to which a human actor feels stimulated to act or grow to greater activity when using a source/piece of information or an IS/IT artifact.

Appendix E: Phase 2 Delphi questionnaire

Phase 2 Delphi Questionnaire

1) How do you go about generating design ideas for a new product? Are there phases that you systematically use? For example, do you use a list of known features you'd like to generate? Or do you use some kind of Strength, Weakness, Opportunities, Threats (SWOT) analysis? (If you have a step by step method when designing an IT application, please list the steps you use. If not, and you have an ideal of what you would do, feel free to provide those steps.)

2) Do you do anything different in your design method when modifying an existing application versus designing a new application?

3) If you have ever done A/B Experimentation, what are unique things to consider when designing treatment versions of the pages to use for the Experiment.

In this study, there are 10 iOS web applications. Each application was evaluated by trained raters using a 1 to 5 Likert scale in 8 unique scenarios on 12 attributes from the TEDS framework. Each evaluation also has some number of screenshots and comments correlated to the ratings which were submitted by the evaluators as they completed their evaluations. The comparative ratings are available in detail and aggregate in Excel using conditional formatting showing how each application compares to the other applications in each scenario and on each of the 12 TEDS attributes. The correlated screenshots and comments are also available to see using a tool which can search and filter based on Scenario, Attribute, and Application.
The purpose of this focus group is to brainstorm a method to use the evaluation data to create feature improvement specifications which would describe new treatments of one of the applications to be used in A/B Testing.

4) Given what you know about this data, and the goal of this focus group, what steps would you use if you were responsible to design these improvement specifications?

Appendix F: Compiled Phase 2 Questionnaire results

Related	Question	Type	Idea
A	1	Method	Each project is different. But generally speaking, design ideas emerge from various user studies (e.g., contextual inquiry, semi-structured interviews, focus group, co-design, etc.). For example, "future workshops" is a participatory design method that I used in one of my projects. It comprises of four phases: critique, fantasy, and implementation.
A	1	Idea Generation	The critique phase is a brainstorming session on problems of the current situation. Contributions are formulated as a brief, critical statements. No justification is needed.
A	1	Idea Generation	The fantasy phase is oriented toward unrestricted ideas on what the future situation could be like. Participants shouldn't worry about practical and technical limitation. No ideas are considered too extreme. All criticism and judgment of the feasibility is postponed during this session. Ideas are clustered and refined into "visions." Some researchers recommend using metaphors as a way to summarize and develop the vision.
A	1	Idea Generation	The implementation phase focus on the possibility of realizing visions under current conditions. This includes identifying what needs to be changed, what needs to be done, when, and by whom.
	1	Tool	Attempt to describe the feature in a spec with pictures, workflows, data models, etc. Should include a listing of features and benefits
	1	Tool	Use design patterns from another application that might provide a framework
B	1	Idea Generation	Observing customers: - live - focus groups or just using product - feedback - bugs, tickets, social media - data – observing logs, analytics around product usage to identify opportunities
B	1	Idea Generation	Competitive research (weakness/threat part of SWOT analysis)

B	1	Idea Generation	Opportunities based on longer range planning and strategic direction
C	1	Idea Generation	Data wallow … competitive analysis, product analysis, focus group feedback ... etc
C	1	Idea Generation	Come up with a north star statement … What are we trying to do
C	1	Idea Generation	Consider user goals (what do they want/need) and business goals (what do the product people want/need)
C	1	Idea Generation	Start with scenarios we want to enable – experiences we want to create, and build off those to feature list (sometimes include scenarios we are explicitly not addressing)
C	1	Tool	Group sketch sessions – have folks working together on product, sketch out ideas – offering appropriate canvases as base.
C	1	Tool	Share ideas – pull out the good idea nuggets
C	1	Tool	Wireframe (content and navigation framework), prototype, create comps and UX flow
C	1	Tool	Create image boards for visual inspiration
C	1	Idea Generation	Get user feedback on ideas, and iterate.
	1	Tool	Look for designs that are familiar with other apps (design patterns)
D	2	Tool	Start with original screenshots of the appliciation and tweak
D	2	Tool	Make alterations to the pages with new buttons, layout, etc, reflecting the new functionality
E	2	Idea Generation	Designing a new application would get more time and attention, for example possibly trialing in a forum/focus group, compared to just modifying something.
E	2	Tool	When modifying something that already exists, presumably the scope is narrower to fix or improve a known issue and changes should ideally be done in small increments and measured or AB tested along the way. In both cases, UX workflows are mapped out on whiteboards or wireframes as a key step.
F	2	Tool	Modifying an existing application generally means more existing design constraints. There are standard conventions, guidelines, and requirements to follow.

F	2	Tool	Designing a new application, on the other hand, is all about thinking outside the box.
G	2	Idea Generation	UX Research: Get user feedback; on product pain points, what might be missing. Collaborative design with users
G	2	Tool	Data metrics: Collect any available data to see if it informs the design, in regard to what people actually use, what they go elsewhere for.
G	2	Tool	Is benchmark testing appropriate (so you have a baseline of if you've improved what you set out to improve)
G	2	Idea Generation	You actually have a user base – involve them
G	2	Tool	You actually have code – is it easy to modify to variations/new ideas.
	2	Tool	Modifying an existing application would be more about sourcing information from users about small changes that would make their experience better. Whereas if an app is generally performing poorly in the marketplace a more advanced re-assessment of the requirements and user scenarios may be vital to a re-design.
H	3	Tool	Change one thing at a time so you can isolate specific changes that are helpful/harmful
H	3	Tool	Consider the metrics you will use to evaluate success
H	3	Tool	Try to run as many tests/variations as possible presuming that you won't know what will be best
	3	Tool	Keep the number of changes minimal so that it is easier tell which change is making the difference in the metrics
I	3	Tool	Know what you are testing. Don't have multiple variables
I	3	Tool	Question the data you get back. I find its helpful to have somehow who is great at analyzing the kind of data you get back – it's easy to misread results.
I	3	Tool	There are various modes of A/B testing. Its more personal and involves asking questions, make sure they are not biased. Ask for reasoning with answers.
I	3	Tool	I hope for a natural environment – if its measuring a mobile UX – test/get feedback on a phone
I	3	Tool	Have hypothesis of what you hope to find out, and how you might act on it.

J	4	Tool	I would pivot the ratings by: - Scenario - Cluster - Application
J	4	Tool	If the app is doing well across the board, then I'd focus on: - Comparison in each scenario with the best performing artifact in that scenario and see what values they are doing well with. - Comparison in each cluster to see if there are clusters in which my application does poorly - Comparison in each cluster to see if there is another application which does very well in particular clusters
K	4	Tool	Look at the each #1 rated scenario (regardless of which of the 10 apps it is on). Try to identify what made that scenario "good" and emulate them.
K	4	Tool	Look at the worst rated scenarios, for example which app did the worst job on team news. Study that to understand what makes it bad. Try to avoid those design patterns.
K	4	Tool	Assuming the 12 attributes are important to customers, evaluate your new design against them to consider how it might be rated.
K	4	Tool	Apply those learnings (both good and bad) to my app/scenario that I am trying to build.
K	4	Tool	Run an AB test
K	4	Tool	Invest in those areas that usage data such as logging or analytics shows are weak points/opportunities in the product. Invest less in areas that do not appear to be used, for example if team news is actually not viewed often and not considered to be strategically important, do not invest as much in it.
L	4	Tool	Get clarity on who our intended users are – who are we building this for?
L	4	Idea Generation	What do those people care about – what do they want out of this app. (like/dislike about the site. What do they do elsewhere)?
L	4	Idea Generation	What do they do … when they go to a game, make plans for the season or a game. General 'day in the life' of fan. Who are these people?

L	4	Idea Generation	is there anything we can do on the site that they don't already have (help us invent features)
L	4	Tool	What do the sounders care about?for those areas that the user/sounders rate are super important (that we have data for) … - Which sites do best in those areas.- What is different on those sites, from the poorly rated sites- How can we apply those principles to our specific needs – what would that look like on our site?- Be inspired by the good designs – not necessarily copy them. Understand why they are better.- Illustrate the variation to generate ideas – and then low fi proto to test out.
K	4	Tool	Establish rater perspective/experience with the app/app design - Who are these people and what are they trying to get out of my app? - What is the persona they are using to establish their perspective?
K	4	Tool	Identify trouble spots: lowest ratings - Determine if they are more to do with simple fixes or a larger re-design

Appendix G: Sounders FC Player Information Feature Improvement Treatment Specification

1 Player Information Feature Improvement Spec

Feature improvement specification for the Sounders iOS app in the Player Information Scenario

This spec provides competitive analysis of 11 iOS sports applications for the Player Information scenario based on assessment and evaluation data from a TEDS analysis done in the fall of 2016. The document is organized as follows:

There is a summary table of the ratings of all applications for each of the rated criteria in this study. Next for each criteria, there is a detailed discussion of the applications performed well with screenshot examples from these applications. This discussion is provided to suggest improvement ideas for the Sounders FC application in order to create treatments for A/B Testing.

Contents

Player Information – Summary Table for all Applications for the 12 Criteria

Values	Seahawks	Barca	Sounders Legacy	Real Madrid	Chicago Fire	Bayern	OneFootball	ESPN FC	Brooklyn Nets	Arsenal FC	Sounders FC	Std Dev	Average
Navigation and Findability (Ease of Use) Browsing/searchability, Mediation, Orientation, Simplicity	3.65	3.70	3.10	3.70	3.20	3.50	3.90	1.90	3.80	3.50	4.00	0.58	3.45
Structure (Ease of Use) Formatting/Presentation, Order, Accessibility	3.60	4.00	3.10	3.80	3.10	3.00	3.90	2.30	3.90	3.50	3.60	0.51	3.44
Identity (Noise Reduction) Item Identification, Subject description, Summary, Selectivity, Precision	3.64	3.60	2.90	4.00	3.40	3.00	3.80	2.40	3.70	3.40	3.50	0.46	3.39
Parsimony (Noise Reduction) Linkage/Referral, Order, Novelty	3.33	3.80	2.90	3.50	2.90	2.60	3.60	2.10	2.80	3.00	3.60	0.51	3.10
Completeness (Quality) Accuracy, Comprehensiveness, Currency	3.87	3.40	1.70	4.00	2.90	3.40	3.90	2.40	3.68	3.40	3.20	0.70	3.26
Trustworthiness (Quality) Reliability, Validity, Authority	4.67	4.20	2.80	3.90	3.40	3.70	3.80	2.30	3.60	3.70	3.70	0.63	3.62
Interaction (Adaptability) Contextuality, Transaction, Feedback, Community/Social networking	2.70	3.70	1.80	3.80	1.90	2.90	3.60	1.80	2.50	2.70	3.30	0.75	2.79
Customization (Adaptability) Flexibility, Simplicity, Trust, Individualization, Localization, Privacy	2.93	2.80	2.20	3.10	2.30	2.40	3.30	2.00	2.30	2.20	2.60	0.42	2.56
Savings (Performanc) Cost savings, Time savings	3.20	3.20	2.78	3.80	2.60	3.00	3.20	2.00	3.60	2.70	3.00	0.49	3.01
Confidence (Performance) Security, Safety	3.30	3.30	2.90	3.40	2.60	3.40	3.40	2.20	3.30	3.20	3.30	0.39	3.12
Attractiveness (Affection) Aesthetics, Satisfaction/rewarding/ incenting	3.30	3.80	2.60	4.40	2.60	2.80	3.70	2.40	3.70	3.30	3.20	0.62	3.25
Enjoyment (Affection) Entertainment, Engagement, Stimulation	3.27	3.40	2.30	4.10	2.60	2.80	3.60	2.20	3.40	3.00	3.20	0.57	3.08
Total	41.46	42.90	31.08	45.50	33.50	36.50	43.70	26.00	40.28	37.60	40.20	5.91	38.07
Average Likert	3.45	3.58	2.59	3.79	2.79	3.04	3.64	2.17	3.36	3.13	3.35	0.49	3.17

Navigation and Findability

Definitions (Navigation and Findability)

Definition - Navigation and Findability (Ease of Use):

How easy is it to find the information you are looking for? Can you easily browse or search? Does the app provide filters to narrow down your search? This ranking refers to the information artifacts' effectiveness evaluated on the aspects of:

- ***Browsing/browsability/searchability*** - Human actor-expected capability of an IS/IT artifact to scan an information neighborhood with the probability that the human actor will serendipitously find information of value. (Taylor, 1986, p. 70)
- ***Orientation*** - The means used to assist human actors in getting answers from the IS/IT artifact. (Taylor, 1986, p. 70). The act or process of intervening between the IS/IT artifact and the human actor to promote reconciliation or understanding, that is, the intermediation between the functions and characteristics of the IS/IT artifact and the human actor.
- ***Mediation*** - The means used to help human actors understand and to gain experience with the IS/IT artifact and its complexities. (Taylor, 1986, p. 70). The ease with which a human actor can orient himself/herself regarding the intended or emerging utilization of the IS/IT artifact at hand; that is, the IS/IT artifact gives clear clues to where the human actor can find information or functions such as intuitive navigation, breadcrumbs, and other visual and audio clues.
- ***Simplicity*** - Lack of complexity, complication, or difficulty when a human actor interacts with and operates an IS/IT artifact.

Ratings – Navigation and Findability:

Values	Seahawks	Barca	Sounders Legacy	Real Madrid	Chicago Fire	Bayern	OneFootball	ESPN FC	Brooklyn Nets	Arsenal FC	Sounders FC	Std Dev	Average
Navigation and Findability (Ease of Use) Browsing/searchability, Mediation, Orientation, Simplicity	3.65	3.70	3.10	3.70	3.20	3.50	3.90	**1.90**	3.80	3.50	4.00	0.58	3.45

Highest Rated: Sounders FC

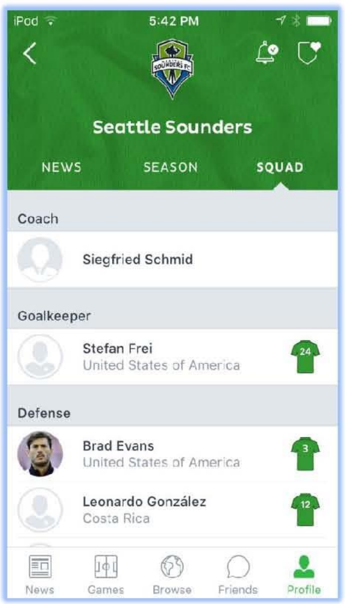
Other highly rated apps: in Navigation and Findability for Player Info include: One Football (3.9) and Brooklyn Nets (3.8). Arsenal and Bayern also had positive feedback from raters in the screenshots and comments.

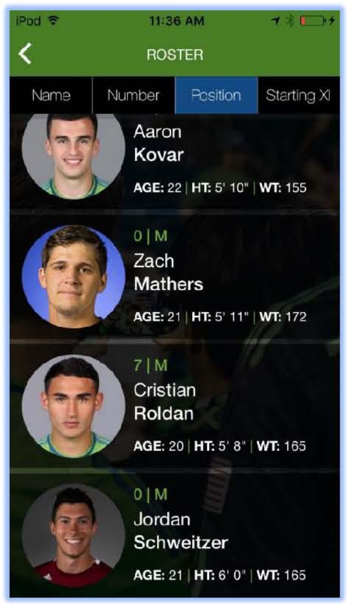
Evaluations (Navigation and Findability)

One Football

Comments from raters, sample screenshots and discussion:

- Took me a bit to figure out how to find players, but once I did the metaphor makes sense. You have to drill down to a team to find players first, but once there it looks great.
- I haven't figured out a way to search for any player in soccer without making their team a favorite first. Then searching for them on that team.
- There's no searching, filtering, or changing the order of players on a roster.
- Roster is organized by position. Coach is included in the list
- Can't reduce the positions in the list as you can with Arsenal. So you have to scroll through the whole list.

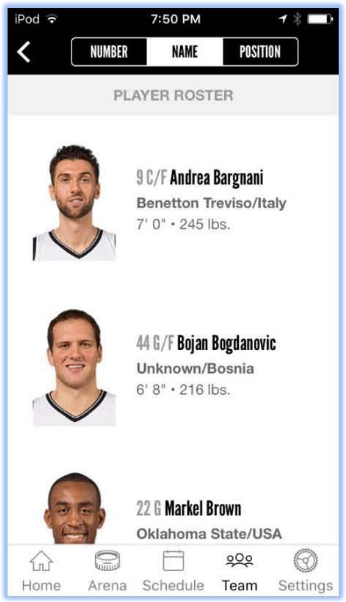


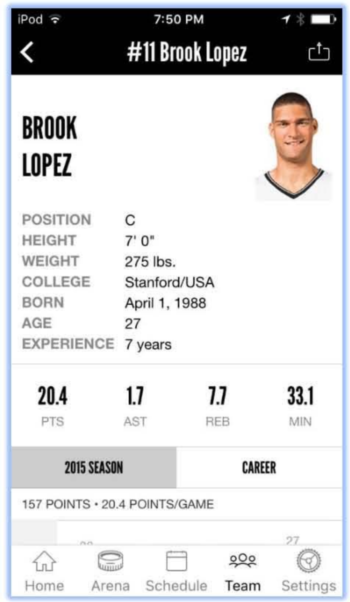


Brooklyn Nets

Comments from raters, sample screenshots and discussion:

- Good use of filters. Almost like how the Seahawks did it!
- The “Team” menu item is a little hidden at first at the bottom. Then on the “Team” page there is a link to the “Roster”
- The roster isn't easy to find. It's off of the Team menu option, but the default team page is not the roster it's a basic summary list of stats and upcoming games. There is a link to Roster, which isn't clear that it's clickable.
- List can be ordered by Number, Name and Position.
- Can’t filter the roster list.
- The order by position doesn’t break up the positions with any delimiter signal in the list like some of the other apps.

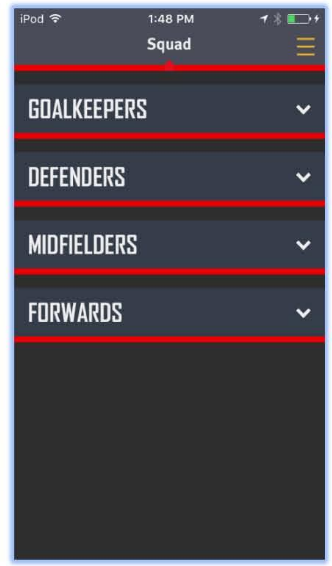




Arsenal:

Comments from raters, sample screenshots and discussion:

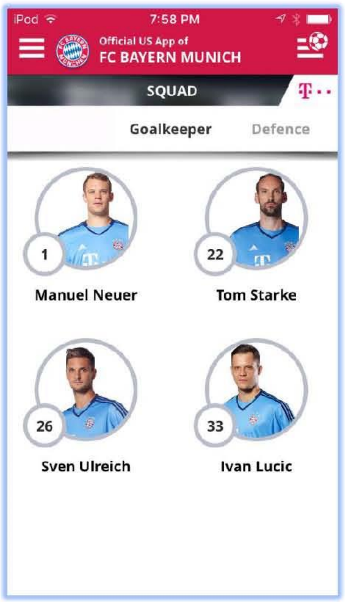
- The Player Information is found off of the Roster menu item.
- There's also a "Team Info" menu item that shows the injured players.
- Clicking on one of them in the list also brings up the player card.
- The list of players is divided by position. That is the only sorting allowed on the page for players. (Sounders FC has 4 ways to sort – Name, Number, Position, Starting XI)
- Each position group can be expanded or compressed clicking on an icon. This is a nice feature to quickly reduce the length of the list and break things up. (Perhaps something the Sounders FC app could consider when the list is ordered by "Position".)
- There is no search. Filtering is done with the compressing/expanding on the position group in the list.

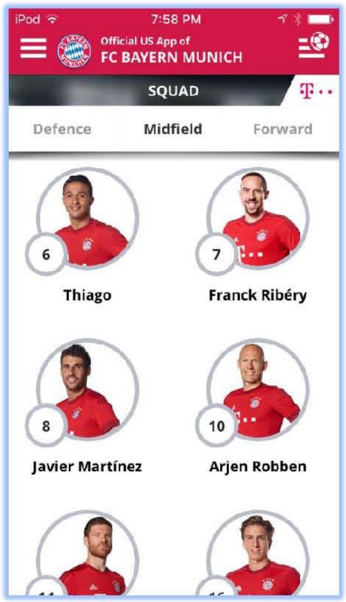


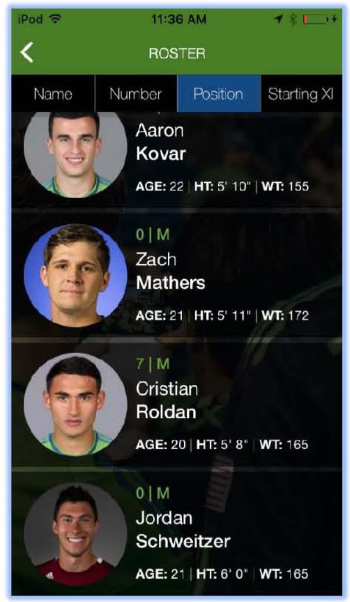
Bayern Munich

Comments from raters, sample screenshots and discussion:

- No search, but the filtering by position with good sized photos makes this effective
- Notice that the photos of the players aren't cut off. Sounders have some headshots for players that are cut off.
- The consistent background color of white is good. Sounders photos seem to switch for some players, and the color uniform also changes.
- Sounders has more meta data (age, height, weight) in the listing. But at first glance it's wasn't obvious to me the "M" meant midfield, nor that the number listed was the number. Is there a way to make that information pop more, or become more evident.







Real Madrid

Comments from raters, sample screenshots and discussion:

- Not clear that the findability is something great here. They have a scrolling page which allows you to flip through the players, but apparently there is no roster listing.

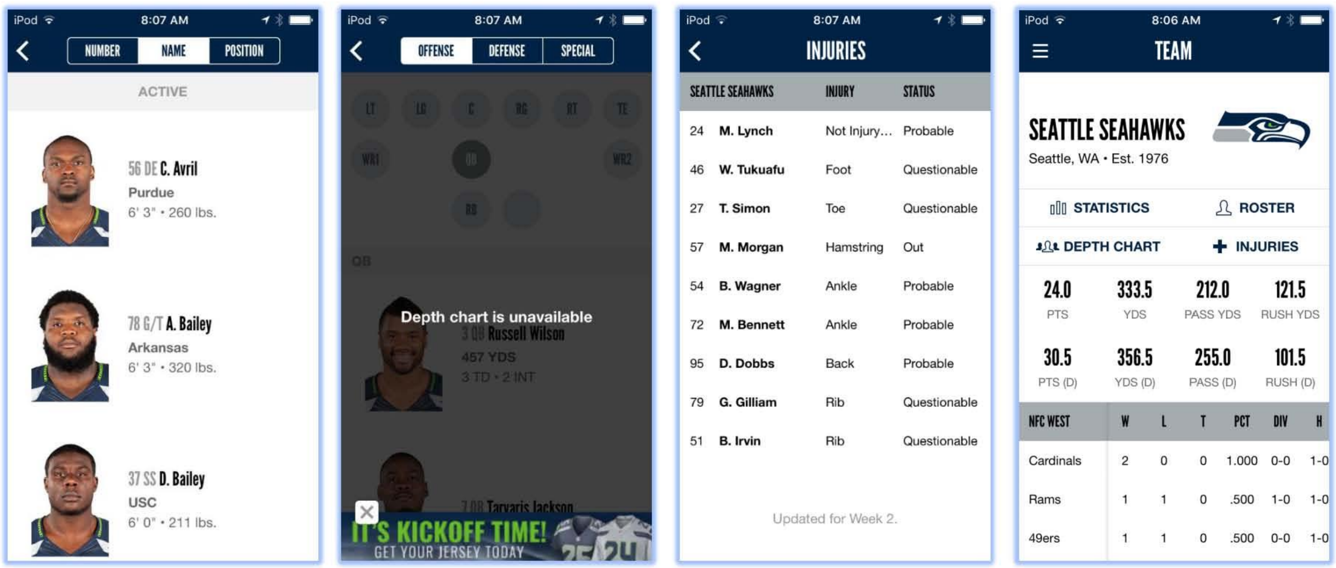
- The roster listing is a scrolling left to right at the bottom of the page, with the currently selected player on the top. Kind of interesting UI.
- Allows for quick view of players. Consistent pose, uniform, background on this page. It looks really clean with the somewhat see-through logo in the background.
- No resorting
- Only includes 22 players in the list. Can’t sort them. Can’t filter them. But the treatment is really nice. Super clear photos. Each player is posed in consistent positions which makes the scrolling through the list less jarring.
- It’s an interesting treatment to have the highlighted player take the upper 70% of the page.
- There is plenty of room for more metadata about the player on this scrolling list for the currently selected player besides just the jersey number. But it makes the list look very clean.

Seattle Seahawks

Comments from raters, sample screenshots and discussion:

Browse-able yes. There are 4 ways to navigate the player info from the “Team” page. The Team page leads to the kind of Roster page that most teams have.

- Statistics
- Roster
- Depth Chart
- Injuries



All of these views are allow for some browsing. Currently the depth chart is down, but I did see it earlier and it's useful to see the positions each player is at. Could be educational for folks as well. Even the layout of positions that the Sounders play would be useful. Notice on the Seahawks page, even though it's greyed out in the example that the positions are there.

Possible to add a "depth chart" for the Sounders? Could be useful feature and education for the Sounders.

Possible to add an "Injuries" tab?

Possible Sounders FC Treatments (Navigation and Findability)

Assumptions/Constraints: Assume that the app doesn't control the photos that are displayed.

Info Display choices (what could we use instead in terms of controls, or graphic choices):
One Football

- One Football uses a grey delimiter between the different positions in the roster list (Goalkeeper, Defense, etc.)
- One Football includes the coach in the Roster page. This is something the Sounders could include as well. If this was the case however it might not make sense to include the coach in different positions in the list based on the sorting order. For example, the coach should likely be at the top or bottom of the list even if the Sounders list is ordered by alphabet on the Name.
- One Football includes a navigation menu at the bottom of all pages (News, Games, Browse, Friends, and Profile). Sounders only include a back arrow '<' as a navigation tool. Sounders could consider including the menu on their pages. It is nicer for navigation.

Brooklyn Nets:

- Like the One Football app, the Brooklyn Nets app includes a navigation menu at the bottom of all pages (Home, Arena, Schedule, Team and Settings). Sounders only include a back arrow '<' as a navigation tool. Sounders could consider including the menu on their pages. It is nicer for navigation.
- Brooklyn Nets has an overall team landing page which allows users to look at either the roster list, or at the team stats. The sounders does not have any team stats option. This is something that could be built and integrated if the data is available.
- Brooklyn Nets have tabs on the player card to allow toggling between career stats and season stats. Sounders could consider doing this if they have access to the data.

Arsenal:

- The Arsenal app has the ability to collapse the roster list subsections (Goalkeepers, Defenders, Midfielders, and Forwards). This is something the Sounders could consider when enumerating the list by position.
- The Arsenal app has much bigger headshots of the players with room to still add the metadata if chosen to. The implication for the Sounders app is that they would show fewer players per page and require more scrolling to see the whole list. Currently on a small iPod screen 4 players fit on a page. For Arsenal only about 2.

Bayern:

- The Bayern Munich app does not have different sort orders but instead uses the position as a filter. The user can click on the position filter at the top to select which position group, or they can swipe to the left or right to get to the previous or next position group. The Bayern app opts to show less data on the roster listing page, in order to show more players on one page. On the iPod, this translates to almost 6 players on a page versus the 4 players on a page that the Sounders app displays. Sounders could consider using the filtering method rather than the sorting method currently used.
- On the Sounders app, it is not clear what "M", "F", "G" mean in the metadata initially whereas the filters for midfield, forward, goalkeeper and defense become obvious on the Bayern app I think. Sounders app should consider expanding the "M" to "Midfield', "F" to Forward, etc.

Real Madrid:

- The Real Madrid app uses 70% of the page with one player and scrolling of the roster left to right on the bottom. The background and the photos are super clear and professional. Nice use of fonts, colors, logo. Sounders could consider doing something similar by showing a larger photo of one player at a time with scrolling left to right on the bottom. Would be interesting to see if this is favored.

Seahawks:

- Similar to the Brooklyn Nets app, the Seahawks have a team landing page which links to team stats, injuries, and depth chart in addition to the roster page. The Sounders app could consider adding a page for team-wide information if this is available in the backend data. Even more useful if the data that is shown is compared to other MLS teams.
- The Seahawks app has a page to show injured players. This would be useful in the Sounders app
- The Seahawks app has a page to show the depth chart. Would be educational for users to see who is ahead of who on the depth chart if available. Also just showing the formation that the Sounders play would be useful.

A/B implications - Can we A/B test this? Metric or metrics can be used for the Overall Evaluation Criteria (OEC) for the experiment:
Can we track how many players are clicked on? Time on page? Return visits?

Structure

Definitions (Structure)

Definition - Structure (Ease of Use):
How easy is it for you to understand how the information presented is ordered; and, how many interactions does it take to get the information you seek? This ranking refers to the information artifacts' effectiveness evaluated on the aspects of:

• ***Formatting/presentation*** - The physical presentation and arrangement of data/information in ways that allow more efficient scanning and hence extraction of items of interest from the store. (Taylor, 1986, p. 70). Presenting and arranging information to facilitate scanning and selecting (Choo, 2002) on part of a human actor.
• ***Order/consistency*** - The value-added by initially dividing or organizing a body of subject matter by some form of gross ordering, such as alphabetization, or large groupings. (Taylor, 1986, p. 70). Orderly and systematic arrangement of IS/IT artifact components and elements such that a human actor can easily identify information and functionality provided by the IS/IT artifact. Also, the fashion through which the IS/IT artifact transforms and presents itself based on human action.
• ***Accessibility*** - The processes of making access to information stores easier in a physical sense. (Taylor, 1986, p. 70). Ease of physical access to an IS/IT artifact and to the information and functionality that the IS/IT artifact can provide to human actors (including to impaired human actors). Also, how many, or how few, steps does the human actor need (for example, in terms of clicks) to reach the desired information or functionality.

Ratings – Structure:

Values	Seahawks	Barca	Sounders Legacy	Real Madrid	Chicago Fire	Bayern	OneFootball	ESPN FC	Brooklyn Nets	Arsenal FC	Sounders FC	Std Dev	Average
Structure (Ease of Use) Formatting/Presentation, Order, Accessibility	3.60	**4.00**	3.10	3.80	3.10	3.00	3.90	**2.30**	3.90	3.50	3.60	0.51	3.44

Highest Rated: FC Barcelona
Other high rated apps: One Football, Brooklyn Nets

Evaluations (Structure)

FC Barcelona

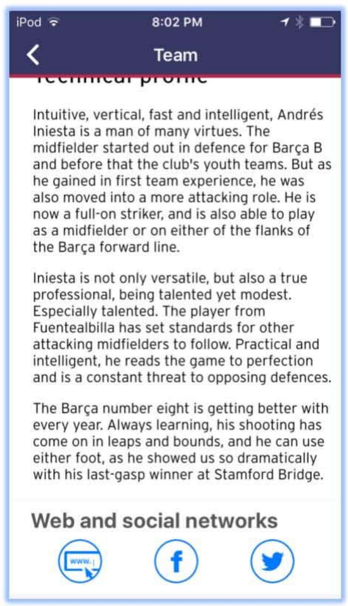
Comments from raters, sample screenshots and discussion:

- The player card looks nice. Clean. It's accessible (few clicks). The presentation and graphics are clean and nice.

- Player picture is cut off in landscape mode. Otherwise, very easy to get what you want and well formatted.
- Two tabs on the player card (Technical card and Stats):

Technical card:

- Demographic data (height, weight, birthday, birthplace)
- Static text describing player
- Links to players twitter feed, player's Facebook page and players own website. The real Icon's used here are helpful to let user know what they are clicking on. If the player does not have these social media outlets available, then the icons are disabled and grey.
- There is way to navigate directly to the previous or next player in the roster list on the player card. Sounders do not have this. You need to go back to the roster list.



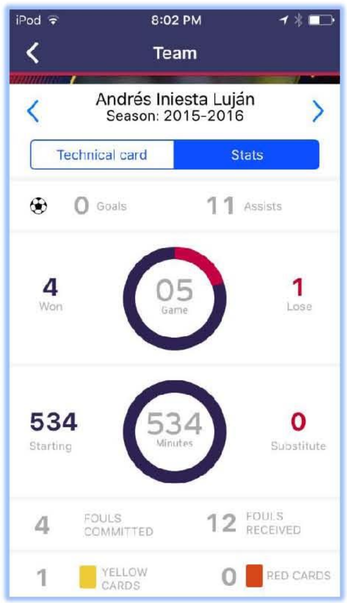
Stats:

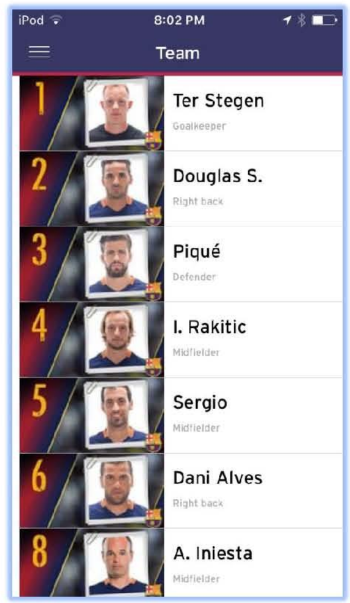
Not an overwhelming amount of stats. Only 5 basic categories. And only using current season stats.

- Goals/Assists
- Wins/Losses in games where player has played
- Minutes played as a starter, and minutes played as a substitute
- Fouls committed and Fouls received
- Yellow cards and Red Cards

Uses circle graph for wins/losses and starting/substitute. Might be interesting visual for these stats.

Sounders might be able to use this for the stats they display. Currently the information is just displayed as text. It's also all career stats, not individual games. Is it possible for Sounders to show both? Can we add an additional tab to show stats?





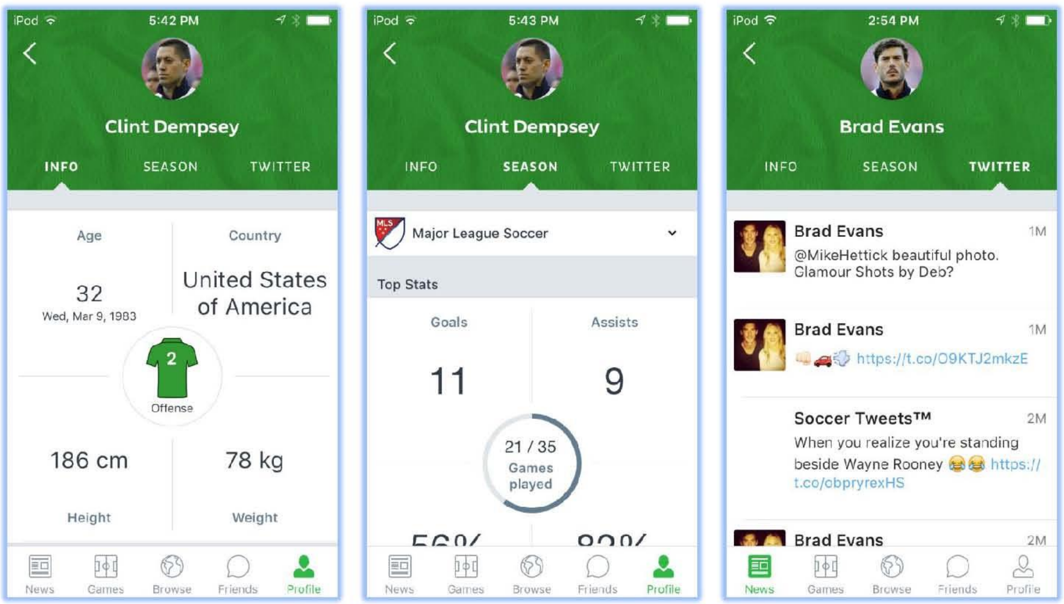
One Football

Comments from raters, sample screenshots and discussion:

- I like the 3 sections on the player card:

- Info
- Season
- Twitter

- Info on the Information Tab is very basic (height/weight/number, age, country). The Season Tab has good stats and there is a filter to select which competition (MLS, Champions League, etc.). The Twitter Tab is just a link to the players twitter feed,
- Overall style for displaying player information is extremely consistent, regardless of any other teams or leagues.

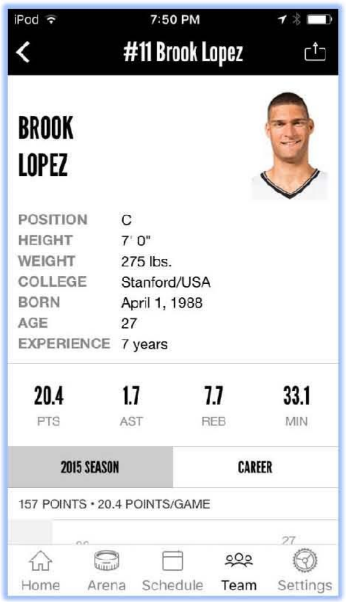


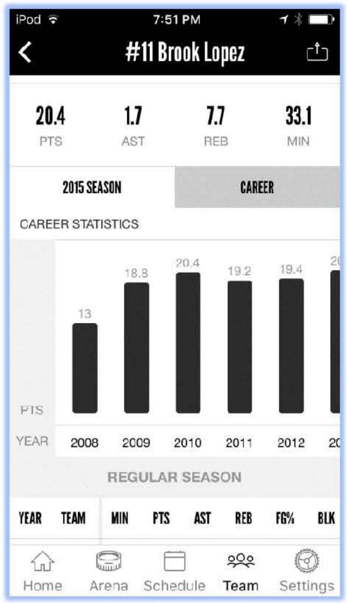
Brooklyn Nets

Comments from raters, sample screenshots and discussion:

- There are 3 sort orders: number, name, position. Can't search, but the roster for a basketball team is not long.
- The layout of the player card is well done. Photo at the top along with demographic information: position, height, weight, college, birthday, age, years in the league. Then summary stats for the player for the current season.

- There is then a filter at the bottom of the page for Current season (2015) and Career. Clicking on the filter will give some really cool stats for the current season versus the Career.
- Interesting, the charts and graphs for the Career has the statistics shown per year, whereas the same graphs in the current season use other teams as the pivot.
- Charts of stats are scrollable left and right as well allowing to see more columns. This isn't obvious in the UI so it is possible many users don't discover.



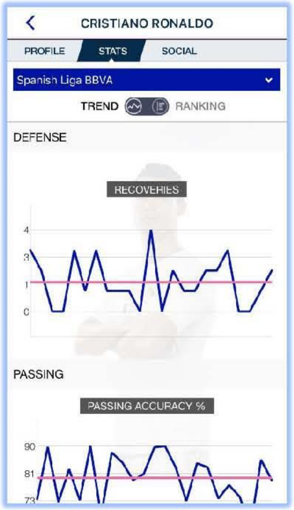


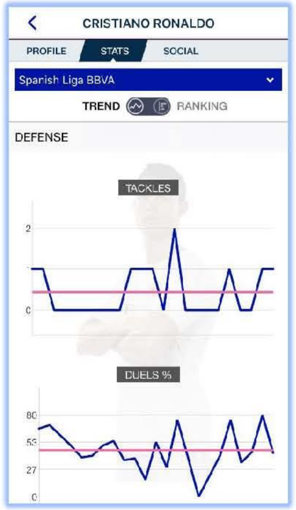
Real Madrid

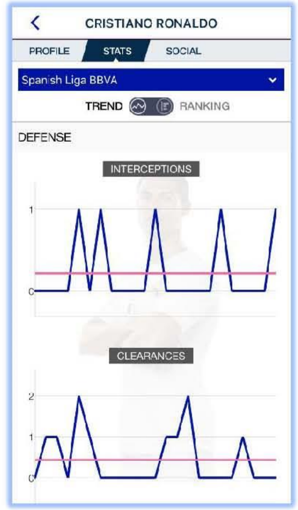
Comments from raters, sample screenshots and discussion:

- The formatting/presentation of the stats is really cool for Real Madrid.
- There's a filter for looking at the stats in different competitions
- There's an option to look at the stats as a trend as well as a ranking among other teammates.

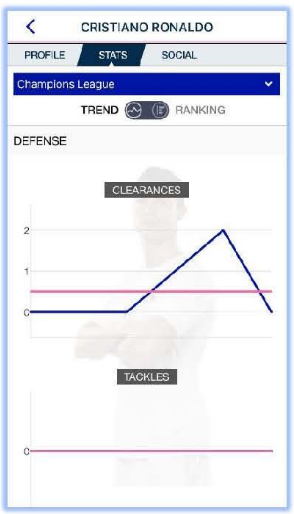

CRISTIANO RONALDO
PROFILE
STATS
SOCIAL
Spanish Liga BBVA
Spanish Liga BBVA
Spanish Copa del Rey
Champions League
PASSES OWN HALF
Cheryshev
Cristiano Ronaldo
Kroos
+ VIEW MORE
PASSING ACC. OPP HALF
Borja Mayoral
Cristiano Ronaldo
Modric
TREND
RANKING
OFFENSE
Arbeloa
Modric
SHOTS %
Casemiro
Lucas Vázquez
GOALS OVER SHOTS %
PASSING
CROSSING ACCURACY %
Benzema
Pepe
TOTAL DRIBBLES
DRIBBLES SUCCESS %
ASSISTS
Shooting Accuracy %
SHOTS ON GOAL
GOAL ASSIST
CHANCES CREATED
CROSSES
PASSES


CRISTIANO RONALDO
PROFILE
STATS
SOCIAL
Spanish Liga BBVA
TREND
RANKING
DEFENSE
RECOVERIES
PASSING
PASSING ACCURACY %


CRISTIANO RONALDO
PROFILE
STATS
SOCIAL
Spanish Liga BBVA
TREND
RANKING
DEFENSE
TACKLES
DUELS %


CRISTIANO RONALDO
PROFILE
STATS
SOCIAL
Spanish Liga BBVA
TREND
RANKING
DEFENSE
INTERCEPTIONS
CLEARANCES


CRISTIANO RONALDO
PROFILE
STATS
SOCIAL
Forward
Age: 31
Height: 185.0
Weight: 80.0
Games played
29
Goals
30
Yellow cards
2
Red cards
0


CRISTIANO RONALDO
PROFILE
STATS
SOCIAL
Champions League
TREND
RANKING
DEFENSE
CLEARANCES
TACKLES

Possible Sounders FC Treatments (Structure)

Assumptions/Constraints: Assume that the app doesn't control the photos that are displayed.

Info Display choices (what could we use instead in terms of controls, or graphic choices):
FC Barcelona:

- The FC Barcelona app has an authored section in the "technical profile" of the player giving a description of the skillset of the player. It's like having a scout or something give a positive spin on this player. The sounders app pretty much just enumerates facts about where the player played previously. This editorial might be something that could provide interesting for many fans.
- The FC Barcelona app links to players twitter feed, player's Facebook page and players own website. The Sounders app only provides links to Facebook and Twitter. They should consider having a link to the players own website.
- The FC Barcelona app has a way to navigate directly to the previous or next player in the roster list on the player card by swiping left and right on the player card. Sounders do not have this. You need to go back to the roster list. This is a feature that should be added for improved navigation.
- The FC Barcelona app has 2 tabs for each player card. The first is the "Technical Card" which includes the static text about the player. The "Stats" tab includes graphical representations of a few stats. This is something that the Sounders should consider. Initially this could be done by splitting off the "Career Statistics" section of the player card on a new page/tab. Then the Sounders player card could have a toggle to switch between the static text and the statistics as the Barcelona app currently does. Barcelona currently shows 5 basic categories. And only using current season stats. The Sounders apps already includes these. The circle graph control isn't especially useful, but it looks cool and could easily resonate. Here are the Barcelona stats on their page.
    - Goals/Assists
    - Wins/Losses in games where player has played
    - Minutes played as a starter, and minutes played as a substitute
    - Fouls committed and Fouls received
    - Yellow cards and Red Cards

One Football:

- Like FC Barcelona, the stats page is a separate tab for each player with a circle graph control. This is something the Sounders should consider doing the same thing.
- The One Football app, like a few others have the ability to toggle between competition (MLS, US Open Cup, Club Friendly Games, and CONCACAF Champions league). If the Sounders app has access to this data, it would be much cooler to be able to do the same and add a filter control to the page. The filter should be set to MLS by default for the stats on the player card.
- Moving the Stats to a page also makes it easier to make the text larger for the stats. It potentially gives more credit for having more stats if the user has to scroll down a bit to see.

Brooklyn Nets

- The Brooklyn Nets app does a good job with stats by having a filter for Current season (2015) and Career. Clicking on the filter will give some really cool stats for the current season versus the Career. As mentioned in other sections, if the Sounders have access to season and career stats it would be helpful to show both. This should be done with a toggle to show either.

- The Brooklyn Nets app charts and graphs for the Career has the statistics shown per year, whereas the same graphs in the current season use other teams as the pivot. It's cool. At the point the Sounders app doesn't use any graphics for the stats, nor do they display any career stats. Both of these items should be considered for feature improvements.
- The Brooklyn Nets app has a link to the iOS sharing icon on the player info page. This allows the user to share the page via mail or notes. Perhaps not that cool, but might be a free feature. Should consider whether or not Sounders app should have this feature on numerous pages.

Real Madrid:

- The Real Madrid app stats is cool and they have more stats than the Sounders. Sounders app should consider adding stats to a second tab. The fact that they use graphic controls makes it look even like they have more. Sounders should consider using visual graphs to display rather than just enumerating the numbers. They are generally line charts. There are two ways to view the stats on the "Stats" tab of the player card:
    - Trend – which shows an average trend line
    - Ranking – which shows how the player does relative to others on the team. I suspect this data is not provided in the Sounders feed?

- The Real Madrid app has a filter for stats in different competitions. This is very cool. If the Sounders have stats per competition then adding a filter on the stats would be cool.

- The Real Madrid app shows the following stats:
    - Static stats shown on the player card
        - Games played
        - Goals
        - Yellow Cards
        - Red Cards
    - Trend (graphed)
        - Defense – Interceptions
        - Defense – Clearances
        - Defense – Tackles
        - Defense – Tackles
        - Defense - Recoveries
        - Passing - Passing Accuracy%
        - Passing – Crosses
        - Passing – Passes
        - Passing – Passing Acc. Final third%
        - Passing – Acc. Opp Hal %
        - Passing – Crossing Accuracy%
        - Passing – Goal Assist

- Offense – Chances Created
- Offense – Shooting Accuracy
- Offense – Shots on goal
- Offense – Goals over shots%
- Offense – Assists
- Offense – Total Dribbles
- Offense – Success%

○ Ranking (graphed)

- Defense – Interceptions
- Defense – Clearances
- Defense – Tackles
- Defense – Duels
- Passing – Passing Accuracy%
- Passing – Passes own half
- Passing – Passing Acc. Opp Half
- Passing – Crossing Accuracy
- Offense – Chances Created
- Offense – Shots%
- Offense – Goals over Shots%

A/B implications - Can we A/B test this? Metric or metrics can be used for the Overall Evaluation Criteria (OEC) for the experiment:
Can we track how many players are clicked on? Time on page? Return visits?

Identity

Definitions (Identity)

Definition - Identity (Noise Reduction):
How easy is it to identify items, based on their descriptions or subject categories? This ranking refers to the information artifacts' effectiveness evaluated on the aspects of:

• ***Item Identification*** - The value achieved by the identification of any information chunk or discrete piece of data by systematic physical description and location information (Taylor, 1986, p. 69). A human actor- oriented comprehensive description of the physical and functional characteristics as well as the location of a unit of information or of an IS/IT artifact.
• ***Subject description/classification/controlled vocabulary*** - The provision of a subject description through access points such as index terms, descriptors, and names (Taylor, 1986, p. 70). Human-actor-oriented interpretive layer, which helps describe, arrange into groups, and classify information or physical and functional characteristics of an IS/IT artifact; that also includes that a human actor can easily identify a source/piece of information or an IS/IT artifact characteristic by its classification.

Ratings – Identity:

Values	Seahawks	Barca	Sounders Legacy	Real Madrid	Chicago Fire	Bayern	OneFootball	ESPN FC	Brooklyn Nets	Arsenal FC	Sounders FC	Std Dev	Average
Identity (Noise Reduction) Item Identification, Subject description, Summary, Selectivity, Precision	3.64	3.60	2.90	4.00	3.40	3.00	3.80	2.40	3.70	3.40	3.50	0.46	3.39

Highest Rated: Real Madrid
Other high rated apps: One Football, Brooklyn Nets, Seahawks, Barca

Evaluations (Identity)

Real Madrid

Comments from raters, sample screenshots and discussion:

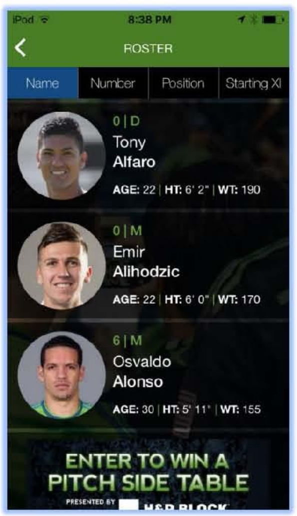
- Big photo of the player
- Stats over the leagues is fairly neat

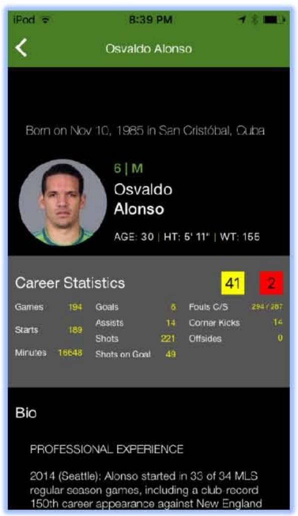
In general the Real Madrid apps uses big enough fonts, pictures, and charts to see and identify the information object. For the player info scenario, it's a focus on the individual players through big photos and colorful stat graphs.


PLAYER CARDS
Select an option
PLAYER CARDS
Don't miss any of them!!
Profile
Stats
Social
TEAM STATS
+250 stats at your fingertips!!
Season
Competition
Match


Player Cards
Select a player
13
Kiko Casilla
Kiko Casilla
Lucas Vázquez
Keylor Navas
Raphael Varane


CRISTIANO RONALDO
PROFILE
STATS
SOCIAL
Forward
Age: 31
Height: 185.0
Weight: 80.0
Games played
29
Goals
30
Yellow cards
2
Red cards
0


iPod
8:38 PM
ROSTER
Name
Number
Position
Starting XI
0 | D
Tony
Alfaro
AGE: 22 | HT: 6' 2" | WT: 190
0 | M
Emir
Alihodzic
AGE: 22 | HT: 6' 0" | WT: 170
6 | M
Osvaldo
Alonso
AGE: 30 | HT: 5' 11" | WT: 155
ENTER TO WIN A
PITCH SIDE TABLE


iPod
8:39 PM
Osvaldo Alonso
Born on Nov 10, 1985 in San Cristóbal, Cuba
6 | M
Osvaldo
Alonso
AGE: 30 | HT: 5' 11" | WT: 155
Career Statistics
41
2
Games
Starts
Minutes
Goals
Assists
Shots
Shots on Goal
Fouls C/S
Corner Kicks
Offsides
Bio
PROFESSIONAL EXPERIENCE
2014 (Seattle): Alonso started in 33 of 34 MLS regular season games, including a club record 150th career appearance against New England

Possible Sounders FC Treatments (Identity)

Assumptions/Constraints: Assume that the app doesn’t control the photos that are displayed.

Info Display choices (what could we use instead in terms of controls, or graphic choices):

Real Madrid:

- The Real Madrid app uses tabs to split up the information across more pages which allows for bigger pictures and fonts. Sounders FC should consider testing this. In particular moving stats to a separate tab.
- The Sounders app could consider increasing the size of the photos and text in the roster list which would reduce the number of players shown at once on the screen, but would help with identifying who and what the user is looking at.

A/B implications - Can we A/B test this? Metric or metrics can be used for the Overall Evaluation Criteria (OEC) for the experiment:
Can we track how many players are clicked on? Time on page? Return visits?

Parsimony

Definitions (Parsimony)

Definition - Parsimony (Noise Reduction):
How well do artifacts provide links to other resources? Is the means to link to those resources novel? This ranking refers to the information artifacts' effectiveness evaluated on the aspects of:

• ***Linkage/referral*** - The value-added when choices are made at the input point of the IS/IT artifact, choices based on the appropriateness and merit of information chunks to the client population served (Taylor, 1986, p. 70). The number and the nature of choices a human actor encounters and chooses from when she or he uses a source/piece of information or an IS/IT artifact.
• ***Order*** - Parsimony, noise-reduced structure, and absence of distraction that a human actor encounters when she or he uses a source/piece of information or an IS/IT artifact.
• ***Novelty*** - The extent of originality and newness relative to a human actors needs that she or he encounters when using a source/piece of information or an IS/IT artifact, that is, the balance of maintaining currency and overloading the human actor.

Ratings – Parsimony:

Values	Seahawks	Barca	Sounders Legacy	Real Madrid	Chicago Fire	Bayern	OneFootball	ESPN FC	Brooklyn Nets	Arsenal FC	Sounders FC	Std Dev	Average
Parsimony (Noise Reduction) Linkage/Referral, Order, Novelty	3.33	**3.80**	2.90	3.50	2.90	**2.60**	**3.60**	**2.10**	2.80	3.00	**3.60**	0.51	3.10

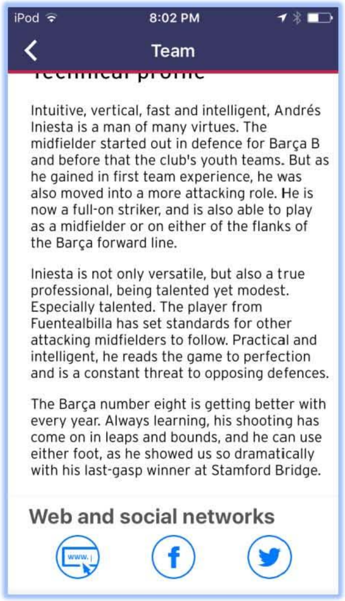
Highest Rated: FC Barcelona
Other high rated apps: One Football, Sounders FC

Evaluations (Parsimony)

FC Barcelona

Comments from raters, sample screenshots and discussion:

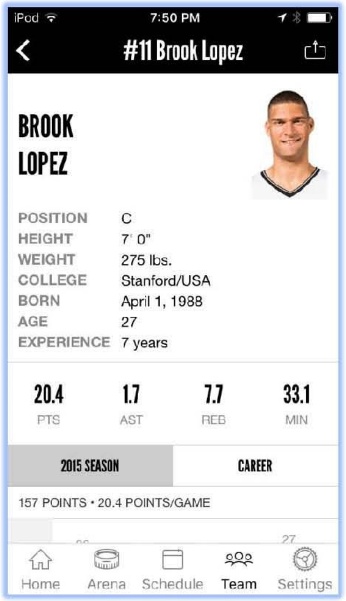
- There is linkage to both the player's home page, and Facebook page if available directly from the player card. This is nice. I think the link is in an iframe too, with a link at the top to get back to the mobile site.



Brooklyn Nets

Comments from raters, sample screenshots and discussion:

- The player cards use filters for "2015 Season" and "Career" which is nice. Also tables to show different stats for playoffs vs regular season.

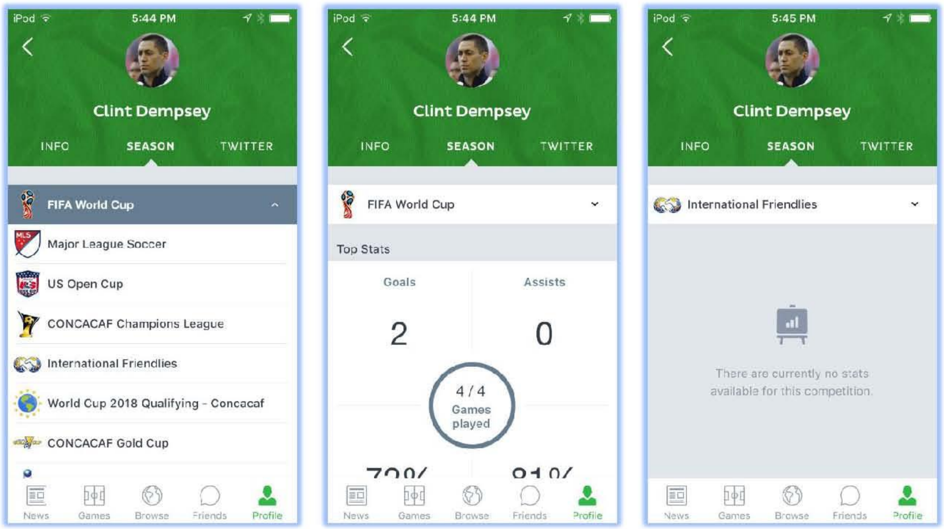


One Football

Comments from raters, sample screenshots and discussion:

- The 3 sections and filter is cool. Allows me to see the stats for a player in different leagues. Only downside is that it appears only some competitions are available. For example Clint Dempsey has MLS and world cup stats, but the other 7 competitions have nothing.

31 Player Information Feature Improvement Spec



Possible Sounders FC Treatments (Parsimony)

Assumptions/Constraints: Assume that the app doesn't control the photos that are displayed.

Info Display choices (what could we use instead in terms of controls, or graphic choices):

Real Madrid:

- Barcelona has links to both Facebook and the players own website when available. The Sounders should consider including these links
- The Brooklyn Nets break out career stats from current year with toggle. This is useful. If the Sounders have access to both yearly stats and career stats this should be done.
- One Football has individual stats for different competitions. This is something the Sounders could also include with a filter if the data is available.

A/B implications - Can we A/B test this? Metric or metrics can be used for the Overall Evaluation Criteria (OEC) for the experiment:
Can we track how many players are clicked on? Time on page? Return visits?

Completeness

Definitions (Completeness)

<u>Definition - Completeness</u> <u>(Quality)</u>
How accurate, comprehensive and current do you find the information offered? This ranking refers to the information artifacts' effectiveness evaluated on the aspects of:

• ***Accuracy*** - The value-added by system processes that assures error-free transfer of data and information as it flows through the IS/IT artifact and is eventually displayed to a human actor (Taylor, 1986, p. 70). Exactness of a piece of information and its conformity to the original source as well as the exact and unaltered transfer and presentation of a piece of information to a human actor through an IS/IT artifact.
• ***Comprehensiveness*** - The value-added by the completeness of coverage of a particular subject or of a particular form of information (Taylor, 1986, p. 70). Complete and broad coverage of a particular subject, in which a human actor is interested, provided by a source/piece of information or an IS/IT artifact.
• ***Currency*** - The value-added (a) by the recency of the data acquired by the system; and (b) by the capability of the IS/IT artifact to reflect current modes of thinking in its structure, organization, and access vocabularies (Taylor, 1986, p. 70). The recency of a source/piece of information sought and the capability of an IS/IT artifact to reflect and represent current modes of understanding of the sought subject matter to a human actor.

Ratings – Completeness:

Values	Seahawks	Barca	Sounders Legacy	Real Madrid	Chicago Fire	Bayern	OneFootball	ESPN FC	Brooklyn Nets	Arsenal FC	Sounders FC	Std Dev	Average
<u>Completeness (Quality)</u> Accuracy, Comprehensiveness, Currency	3.87	3.40	1.70	4.00	2.90	3.40	3.90	2.40	3.68	3.40	3.20	0.70	3.26

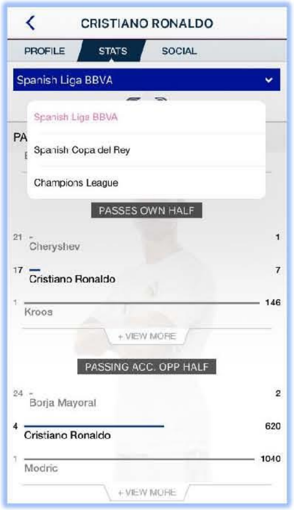
Highest Rated: Real Madrid
Other high rated apps: One Football, Seahawks, Brooklyn Nets

Evaluations (Completeness)

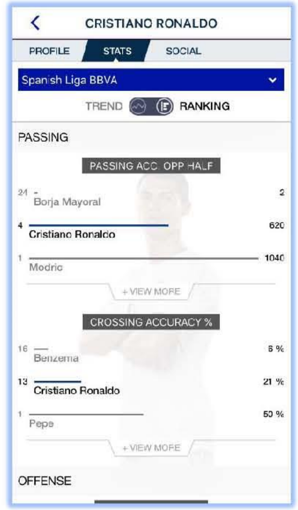
Real Madrid

Comments from raters, sample screenshots and discussion:

- This is the most stats in player info I have ever seen


CRISTIANO RONALDO
PROFILE
STATS
SOCIAL
Spanish Liga BBVA
Spanish Liga BBVA
Spanish Copa del Rey
Champions League
PASSES OWN HALF
Cheryshev
Cristiano Ronaldo
Kroos
+ VIEW MORE
PASSING ACC. OPP HALF
Borja Mayoral
Cristiano Ronaldo
Modric
+ VIEW MORE


CRISTIANO RONALDO
PROFILE
STATS
SOCIAL
Spanish Liga BBVA
TREND
RANKING
OFFENSE
Arbeloa
Cristiano Ronaldo
Modric
+ VIEW MORE
SHOTS %
Casemiro
Cristiano Ronaldo
Lucas Vázquez
+ VIEW MORE
GOALS OVER SHOTS %
Modric


CRISTIANO RONALDO
PROFILE
STATS
SOCIAL
Spanish Liga BBVA
TREND
RANKING
PASSING
PASSING ACC. OPP HALF
Borja Mayoral
Cristiano Ronaldo
Modric
+ VIEW MORE
CROSSING ACCURACY %
Benzema
Cristiano Ronaldo
Pepe
+ VIEW MORE
OFFENSE

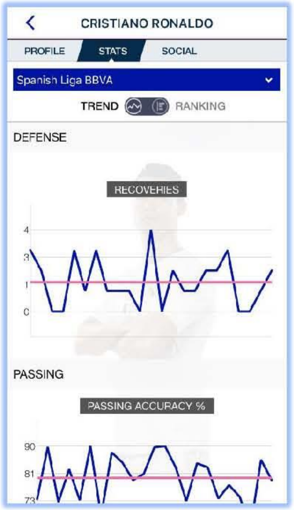

CRISTIANO RONALDO
PROFILE
STATS
SOCIAL
Spanish Liga BBVA
TREND
RANKING
OFFENSE
TOTAL DRIBBLES
DRIBBLES SUCCESS %

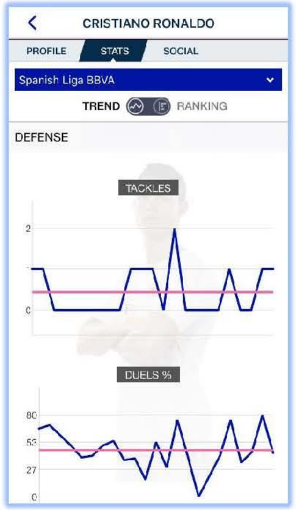

CRISTIANO RONALDO
PROFILE
STATS
SOCIAL
Spanish Liga BBVA
TREND
RANKING
OFFENSE
ASSISTS

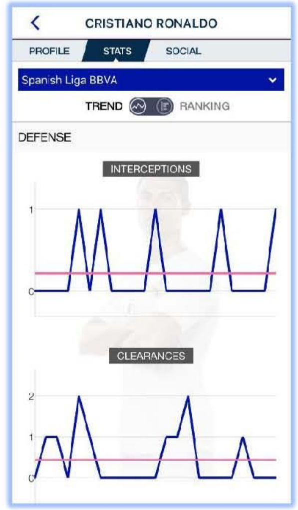

CRISTIANO RONALDO
PROFILE
STATS
SOCIAL
Spanish Liga BBVA
TREND
RANKING
OFFENSE
Shooting Accuracy %
SHOTS ON GOAL


CRISTIANO RONALDO
PROFILE
STATS
SOCIAL
Spanish Liga BBVA
TREND
RANKING
PASSING
GOAL ASSIST
OFFENSE
CHANCES CREATED

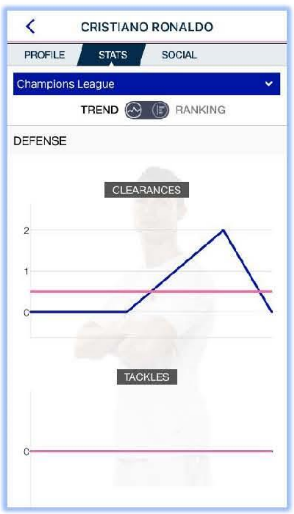

CRISTIANO RONALDO
PROFILE
STATS
SOCIAL
Spanish Liga BBVA
TREND
RANKING
PASSING
CROSSES
PASSES


CRISTIANO RONALDO
PROFILE
STATS
SOCIAL
Spanish Liga BBVA
TREND
RANKING
DEFENSE
RECOVERIES
PASSING
PASSING ACCURACY %


CRISTIANO RONALDO
PROFILE
STATS
SOCIAL
Spanish Liga BBVA
TREND
RANKING
DEFENSE
TACKLES
DUELS %


CRISTIANO RONALDO
PROFILE
STATS
SOCIAL
Spanish Liga BBVA
TREND
RANKING
DEFENSE
INTERCEPTIONS
CLEARANCES


CRISTIANO RONALDO
PROFILE
STATS
SOCIAL
Forward
Age: 31
Height: 185.0
Weight: 80.0
Games played
29
Goals
30
Yellow cards
2
Red cards
0


CRISTIANO RONALDO
PROFILE
STATS
SOCIAL
Champions League
TREND
RANKING
DEFENSE
CLEARANCES
TACKLES

One Football

Comments from raters, sample screenshots and discussion:

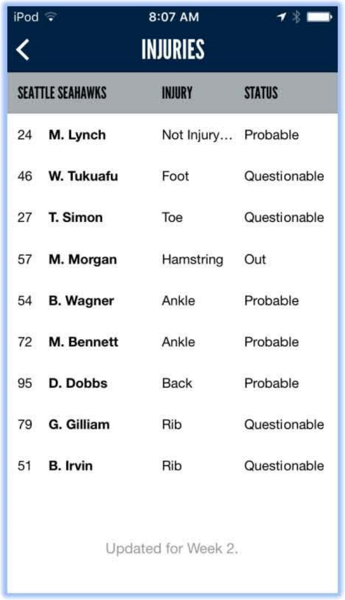
- Giving One Football the benefit of the 4 here because of the stats they have that other sites don't consolidate. There's more personal info on the player card that they could include.

From a statistics perspective the integration of all competitions and career is very complete. However there is really no editorial commentary or personal information included.

Seahawks

Comments from raters, sample screenshots and discussion:

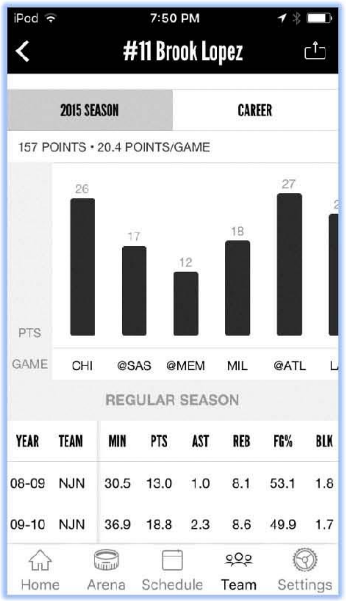
- The injury report is a good example of current information for the Seahawks site. It appears to be current after this most recent game.



Brooklyn Nets

Comments from raters, sample screenshots and discussion:

- Stats are comprehensive. Reminds me of the MLB site for teams.



Possible Sounders FC Treatments (Completeness)

Assumptions/Constraints: Assume that the app doesn't control the photos that are displayed.

Info Display choices (what could we use instead in terms of controls, or graphic choices):
Real Madrid:

- Real Madrid shows an abundance of individual stats with cool graphics. Sounders should look at something they could also do by putting the stats on separate tab (page) and being able to increase the number of stats based on what they already have available.

Seattle Seahawks:

- The Seahawks app has an injury report page that ads to the comprehensiveness of the app. Sounders should include this feature provided that they have access to the injury status. It could be included as a tab off the team page.

Brooklyn Nets:

- The Brooklyn Nets app has individual stats that look very similar to the comprehensive stats shown on a league wide site like MLB. Do the Sounders have access to league wide stats? If so, this might be a place where player stats could be linked to a larger corpus of data.

A/B implications - Can we A/B test this? Metric or metrics can be used for the Overall Evaluation Criteria (OEC) for the experiment:
Can we track how many players are clicked on? Time on page? Return visits?

Trustworthiness

Definitions (Trustworthiness)

Ratings – Trustworthiness:

Values	Seahawks	Barca	Sounders Legacy	Real Madrid	Chicago Fire	Bayern	OneFootball	ESPN FC	Brooklyn Nets	Arsenal FC	Sounders FC	Std Dev	Average
Trustworthiness (Quality) Reliability, Validity, Authority	4.67	4.20	2.80	3.90	3.40	3.70	3.80	2.30	3.60	3.70	3.70	0.63	3.62

Highest Rated: Seahawks
Other high rated apps: FC Barcelona, Real Madrid, One Football

Evaluations (Trustworthiness)

FC Barcelona

Comments from raters, sample screenshots and discussion:

- There is no reason to believe that this data is anything but valid, reliable and provided with authority.
- It looks official. Feels official and professional.

Possible Sounders FC Treatments (Trustworthiness)

Assumptions/Constraints: Assume that the app doesn't control the photos that are displayed.

Info Display choices (what could we use instead in terms of controls, or graphic choices):

- Even though there are differences in the ratings between the different applications, with the Seahawks standing above the rest there are few comments and no screenshots correlated to the rating to suggest why the Seahawks were above average. In general this rating reflects that the raters believe that the data is accurate and not biased. So adding citation, or adding external produced content to show that the information isn't coming merely from the app would increase the trustworthiness.

- If the Sounders wanted to improve the rating here, they could consider adding additional links to external content related to the player and allow comments on the articles.

A/B implications - Can we A/B test this? Metric or metrics can be used for the Overall Evaluation Criteria (OEC) for the experiment:
Can we track how many players are clicked on? Time on page? Return visits?

Interaction

Definitions (Interaction)

Definition - Interaction (Adaptability)

How well does the system provide you with feedback; and, how easily are you able to consume and share articles of interest to your local network? This ranking refers to the information artifacts' effectiveness evaluated on the aspects of:

• ***Contextuality/closeness to problem*** - The value-added by the activities of the system, usually through human intervention, to meet the specific needs of a person in a particular environment with a particular problem; this implies knowledge of that persons style, bias, idiosyncrasies, and sophistication, as well as the politics and constraints of the context (Taylor, 1986, p. 70). The extent to which a source/piece of information or an IS/IT artifact matches a human actors specific informational or transactional needs, also with respect to that human actors specific location when stationary or on the move.

• ***Transaction*** - The capacity of an IS/IT artifact to immediately and on a per-demand basis cope with a human-actors transactional need without referral or deferral (a transaction might include the purchasing of goods and services, electronic or nonelectronic).

• ***Feedback*** - The capacity of a source/piece of information or an IS/IT artifact to entertain, receive feedback from and display feedback to human actors who utilize and assess that source/piece of information or that IS/IT artifact.

• ***Community/social networking*** - The capacity of a source/piece of information or an IS/IT artifact to help human actors form a community or social network, electronically or nonelectronically, around a set of shared interests.

Ratings – Interaction:

Values	Seahawks	Barca	Sounders Legacy	Real Madrid	Chicago Fire	Bayern	OneFootball	ESPN FC	Brooklyn Nets	Arsenal FC	Sounders FC	Std Dev	Average
Interaction (Adaptability) Contextuality, Transaction, Feedback, Community/Social networking	2.70	3.70	1.80	3.80	1.90	2.90	3.60	1.80	2.50	2.70	3.30	0.75	2.79

Highest Rated: Real Madrid

Other high rated apps: FC Barcelona, One Football

Evaluations (Interaction)

Real Madrid

Comments from raters, sample screenshots and discussion:

- Social page integrated into the player card. Very well integrated.
- There's potential to provide feedback via social media on the player's social accounts.

FC Bayern Munich

Comments from raters, sample screenshots and discussion:

- No social networking, but there's a link to buy a jersey from the player card.

Possible Sounders FC Treatments (Interaction)

Assumptions/Constraints: Assume that the app doesn't control the photos that are displayed.

Info Display choices (what could we use instead in terms of controls, or graphic choices):

- The FC Bayern Munich application has a link on the player card to purchase a jersey for the selected player. The sounders application could provide a similar link to take the user to a search results page on the store with the player's available jerseys listed.

- Interaction refers to the ability of the user to interact with the scenario. Social Networking and Feedback are two of the characteristics included in this criteria. Better integration with Twitter where the user could actually post within the app would be one thing the Sounders app could do to improve this scenario instead of just linking off to the site.

A/B implications - Can we A/B test this? Metric or metrics can be used for the Overall Evaluation Criteria (OEC) for the experiment:
Can we track how many players are clicked on? Time on page? Return visits?

Customization

Definitions (Customization)

Definition - Customization (Adaptability)
How well or easily does the app allow you to focus your interactions on local communities of shared interest, whether by using local languages or offering you a flexible means to narrow down content based on relevance to your location? This ranking refers to the information artifacts' effectiveness evaluated on the aspects of:

- ***Flexibility*** - The capability of an IS/IT artifact to provide a variety of ways and approaches of working dynamically with the data/information in a file (Taylor, 1986, p. 70). The dynamic adjustment of information or an IS/IT artifact to a human actors changing informational or transactional needs.
- ***Simplicity*** - The value achieved by presenting the most clear and lucid (explanation, data, hypothesis, or method) among several within quality and validity limits (Taylor, 1986, p. 70). The lack of complication or difficulty with which information or an IS/IT artifact adjust to a human actors changing informational and transactional needs.
- ***Trust*** - An individual human actors willingness to consider a source/piece of information or an IS/IT artifact trustworthy and to act accordingly based on accumulated experience or other clues such as certificates, ratings, or reviews.
- ***Individualization*** - The capacity of a source/piece of information or an IS/IT artifact to adjust to an individual human-actors specific needs. Two forms of individualization are distinguished: (a) Static or basic individualization, for example, based on a human actors preset preferences and selections including push-updates of information; (b) advanced or dynamic individualization where specific human-actors needs are addressed as those that emerge through utilization and relative to the changing contexts over time (including the inference of changes in human-actors preferences, ambiences, or search patterns as well as suggesting the selection of information or the potential utilization of an IS/IT artifact).
- ***Localization*** - The extent to which a source/piece of information or an IS/IT artifact is sensitive to or reflective of differences in physical measures and metrics, time zones, languages, cultural, and other differences in real time relative to a human-actors specific needs.
- ***Privacy*** - A human actors right to be left alone (Warren & Brandeis, 1890) and to be or remain apart from company, observation, tracking, and recording of activities and selections when utilizing a source/ piece of information or an IS/IT artifact.

Ratings – Customization:

Values	Seahawks	Barca	Sounders Legacy	Real Madrid	Chicago Fire	Bayern	OneFootball	ESPN FC	Brooklyn Nets	Arsenal FC	Sounders FC	Std Dev	Average
Customization (Adaptability) Flexibility, Simplicity, Trust, Individualization, Localization, Privacy	2.93	2.80	2.20	**3.10**	2.30	2.40	**3.30**	**2.00**	2.30	2.20	2.60	0.42	2.56

Highest Rated: One Football
Other high rated apps: Real Madrid, Seahawks

Evaluations (Customization)

One Football

Comments from raters, sample screenshots and discussion:

- There is customization in that you can pick favorite teams, and modify over time.

Real Madrid

Comments from raters, sample screenshots and discussion:

- Benefit is that you have to log in to use the app, which is also a negative.

Possible Sounders FC Treatments (Customization)

Assumptions/Constraints: Assume that the app doesn't control the photos that are displayed.

Info Display choices (what could we use instead in terms of controls, or graphic choices):

- Apps such as ESPN and One Football have customization because the app is best suited when the user saves their favorite team which provides a custom list of players for that team. The Sounders don't need this, however there is opportunity to customize a number of other things related to player information when there are more features. Language is one options for greater customization. More flexibility of showing different competitions for a player rather than merely the MLS. More flexibility in showing statistics in different formats. The Sounders FC app could add other language options. As already mentioned in other attributes, there is opportunity to add filters for different competitions, and different formats for stats.

A/B implications - Can we A/B test this? Metric or metrics can be used for the Overall Evaluation Criteria (OEC) for the experiment:
Can we track how many players are clicked on? Time on page? Return visits?

Savings

Definitions (Savings)

Definition - Savings (Performance)
How well is the app structured to save you time in finding what you're interested in, in this scenario? This ranking refers to the information artifacts' effectiveness evaluated on the aspects of:

• ***Time Savings*** - The value achieved by conscious IS/IT artifact design and operating decisions that save dollars for the human actor (Taylor, 1986, p. 70). The extent to which a human actor can cut real cost when utilizing a source/piece of information or an IS/IT artifact.

• ***Cost Savings*** - The perceived value of an IS/IT artifact based on the speed of its response time (Taylor, 1986, p. 70). The extent to which a human actor can save time when utilizing a source/piece of information or an IS/IT artifact.

Ratings – Savings:

Values	Seahawks	Barca	Sounders Legacy	Real Madrid	Chicago Fire	Bayern	OneFootball	ESPN FC	Brooklyn Nets	Arsenal FC	Sounders FC	Std Dev	Average
Savings (Performanc) Cost savings, Time savings	3.20	3.20	2.78	**3.80**	2.60	3.00	3.20	**2.00**	**3.60**	2.70	3.00	0.49	3.01

Highest Rated: Real Madrid
Other high rated apps: Brooklyn Nets

Evaluations (Savings)

Real Madrid

Comments from raters, sample screenshots and discussion:

- Finding the player information is fairly easy.
- Really complete information on the players which makes this app a saver of time from having to go elsewhere for the info.

Brooklyn Nets

Comments from raters, sample screenshots and discussion:

- The stats are great here. Don't need to go elsewhere to dig in and get useful info.

One Football

Comments from raters, sample screenshots and discussion:

- The consolidation of player info across different leagues, and the ability to keep multiple teams in one app is a savings compared to other apps.

Possible Sounders FC Treatments (Savings)

Assumptions/Constraints: Assume that the app doesn't control the photos that are displayed.

Info Display choices (what could we use instead in terms of controls, or graphic choices):

- Savings (time and cost) refer to the ability of the app to meet the user's information needs without having to pay extra money or go to another source to find the information. From a feature perspective, more features and data (provided it's well presented) is better from this perspective. The Sounders app can address this characteristic by increasing the number of stats presented and presenting more detail from other competitions.

A/B implications - Can we A/B test this? Metric or metrics can be used for the Overall Evaluation Criteria (OEC) for the experiment:
Can we track how many players are clicked on? Time on page? Return visits?

Confidence

Definitions (Confidence)

Definition - Confidence (Performance)
How comfortable do you feel using this app? This ranking refers to the information artifacts' effectiveness evaluated on the aspects of:

• ***Security*** - The extent to which a source/piece of information or an IS/IT artifact provides safeguards against fraud and intrusion such that a human actor can feel secure, protected, and free of anxiety when utilizing a source/piece of information or an IS/IT artifact.
• ***Safety*** - The extent to which a source/piece of information or an IS/IT artifact provides safeguards against the risk of hurt, injury, loss, or danger such that a human actor can feel safe when utilizing a source/piece of information or an IS/IT artifact.

Ratings – Confidence:

Values	Seahawks	Barca	Sounders Legacy	Real Madrid	Chicago Fire	Bayern	OneFootball	ESPN FC	Brooklyn Nets	Arsenal FC	Sounders FC	Std Dev	Average
Confidence (Performance) Security, Safety	3.30	3.30	2.90	3.40	2.60	3.40	3.40	2.20	3.30	3.20	3.30	0.39	3.12

Highest Rated: Real Madrid, Bayern Munich, One Football
Other high rated apps:

Evaluations (Confidence)

Not much in the data for the Confidence cluster.

Possible Sounders FC Treatments (Confidence)

Assumptions/Constraints: Assume that the app doesn't control the photos that are displayed.

Info Display choices (what could we use instead in terms of controls, or graphic choices):

- There is really very little distinction with Confidence (Security and Safety) between most of these applications. The comments from the raters refer to providing citation or authority on where data is being sourced. The Sounders application could show citation for where the data is being sourced, either on the page or on an overall "about" page in the settings which also includes the privacy statement.

A/B implications - Can we A/B test this? Metric or metrics can be used for the Overall Evaluation Criteria (OEC) for the experiment:
Can we track how many players are clicked on? Time on page? Return visits?

Attractiveness

Definitions (Attractiveness)

Definition - Attractiveness (Affection)
This ranking refers to the information artifacts' effectiveness evaluated on the aspects of:

• ***Aesthetics*** - The extent to which a human actor appreciates the appearance and perceived beauty of presentation when using a source/piece of information or an IS/IT artifact.
• ***Satisfaction/rewarding/incenting*** - The extent to which a human actor feels personally satisfied, incented, and rewarded when using a source/piece of information or an IS/IT artifact.

Ratings – Attractiveness:

Values	Seahawks	Barca	Sounders Legacy	Real Madrid	Chicago Fire	Bayern	OneFootball	ESPN FC	Brooklyn Nets	Arsenal FC	Sounders FC	Std Dev	Average
Attractiveness (Affection) Aesthetics, Satisfaction/rewarding/ incenting	3.30	3.80	2.60	4.40	2.60	2.80	3.70	2.40	3.70	3.30	3.20	0.62	3.25

Highest Rated: Real Madrid
Other high rated apps: FC Barcelona, One Football, Brooklyn Nets

Evaluations (Attractiveness)

Real Madrid

Comments from raters, sample screenshots and discussion:

- In my opinion, the best looking player info scenario is on this app. Like the colors, the graphics, the photos. Well done

Possible Sounders FC Treatments (Attractiveness)

Assumptions/Constraints: Assume that the app doesn't control the photos that are displayed.

Info Display choices (what could we use instead in terms of controls, or graphic choices):

- The Real Madrid app has very big photos, very clear, big logo and clean consistent backgrounds. The stats are shown in cool graphics. The Sounders app could try using bigger photos. Lighter background. Less text and more visual graphics for the stats. The player card for in the

Real Madrid app is a huge photo with just a few stats. Then the stats are shown on a separate page. Less text on one page. Bring the user to the feature with clean photos, then present the detailed data on a separate page and use graphics to make it more visually appealing.

A/B implications - Can we A/B test this? Metric or metrics can be used for the Overall Evaluation Criteria (OEC) for the experiment:
Can we track how many players are clicked on? Time on page? Return visits?

Enjoyment

Definitions (Enjoyment)

Definition - Enjoyment (Affection)
How engaging and stimulating did you find the app, in this scenario? This ranking refers to the information artifacts' effectiveness evaluated on the aspects of:

• ***Entertainment*** - The extent to which a human actor appreciates the perceived amusement and diversion of presentation and interaction when using a source/piece of information or an IS/IT artifact.
• ***Engagement*** - The extent to which a human actor appreciates the attractiveness of presentation and the appeal of interaction when using a source/piece of information or an IS/IT artifact.
• ***Stimulation*** - The extent to which a human actor feels stimulated to act or grow to greater activity when using a source/piece of information or an IS/IT artifact.

Ratings – Enjoyment:

Values	Seahawks	Barca	Sounders Legacy	Real Madrid	Chicago Fire	Bayern	OneFootball	ESPN FC	Brooklyn Nets	Arsenal FC	Sounders FC	Std Dev	Average
Enjoyment (Affection) Entertainment, Engagement, Stimulation	3.27	3.40	2.30	4.10	2.60	2.80	3.60	2.20	3.40	3.00	3.20	0.57	3.08

Highest Rated: Real Madrid
Other high rated apps: One Football, FC Barcelona, Brooklyn Nets

Evaluations (Enjoyment)

Seahawks

Comments from raters, sample screenshots and discussion:

- The statistics which shows team leaders is nice. The player names are links to the player page. It might be cool to also have NFL wide statistics incorporated here. But that might be tough to pull off.
- The variety of news media allows for engagement with the senses, which in turn can be quite stimulating to the user.

Possible Sounders FC Treatments (Enjoyment)

Assumptions/Constraints: Assume that the app doesn't control the photos that are displayed.

Info Display choices (what could we use instead in terms of controls, or graphic choices):

- Entertainment, Stimulation and Enjoyment make up the Enjoyment cluster. Comments from the raters included a request for more video and pictures. Things that pull the user to look deeper and spend more time. The Sounders app could provide more individual photos, better presentation of statistics. More comparison of how a player performs on various pivots (games, others, leagues).

A/B implications - Can we A/B test this? Metric or metrics can be used for the Overall Evaluation Criteria (OEC) for the experiment:
Can we track how many players are clicked on? Time on page? Return visits?

Recommended Treatments:

	Pri	Title	Description
1	0	Adding Tabs/New Page for Stats on Player Card	Replace 1 scrolling player page, with “Personal” and “Stats” tabs to reduce scrolling
2	0	Bigger Photos, Scroll left and right with Roster on Bottom of page	Replace vertical scrolling roster with small headshots of players, with a large photo of currently selected player and horizontal scrolling through roster.
3	0	Add graphs for stats rather than static text	Use graphs, charts for player stats
4	2	Link to Jersey from Player Card	Provide a link to the store where with the player jersey search terms included

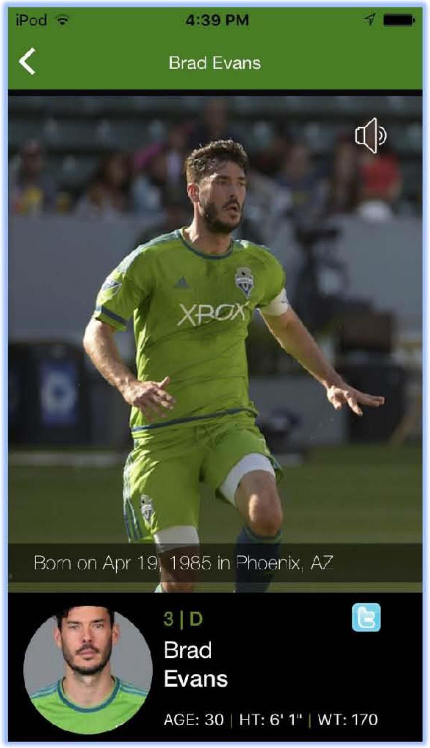
1) Adding Tabs/New Page for Stats

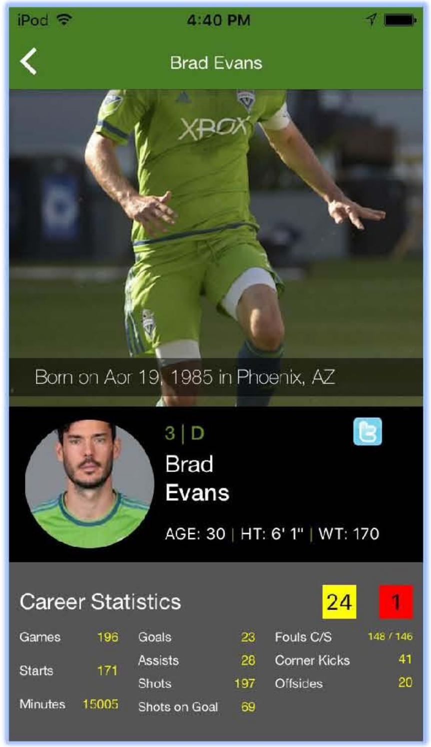
Sounders Current Player Card:

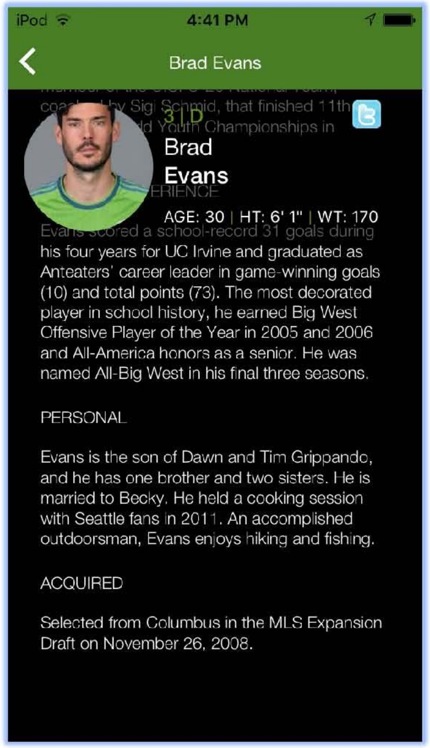
The current implementation has one scrolling page for the Player Card. These 4 screenshots are all of the same page, but scrolled down. Notice how the “Career Statistics are just a grey box embedded? It’s straight text. The suggestion is to move the stats to a different page through a tab, or scrolling. Break up the page. The tabs/toggle is more discoverable than swiping from using the other apps based on the rater feedback.

Examples of this kind of design idea for the player card can be found with the apps from:

- **FC Barcelona** - Tabs for “Technical Card” and “Stats” just below the photo (and filter for which competition).
- **One Football** – Menu options for “Info”, “Season” and “Twitter”.
- **Brooklyn Nets** - Tabs on the player card to allow toggling between “career” stats and “season” stats.
- **Arsenal** - The additional pages scroll to the right. You can see the indicators at the bottom of the page. It’s not very discoverable however compared to the tab/toggle on the page itself.
- **Real Madrid** -The tabs listed at the top of the page: “Profile”, “Stats”, and “Social”
- **Seahawks** – A toggle between “2015 Season” and “Career”. Clicking on this toggles between the current season stats and the career stats. The top of the page stays static, while the bottom changes based on which tab is selected.

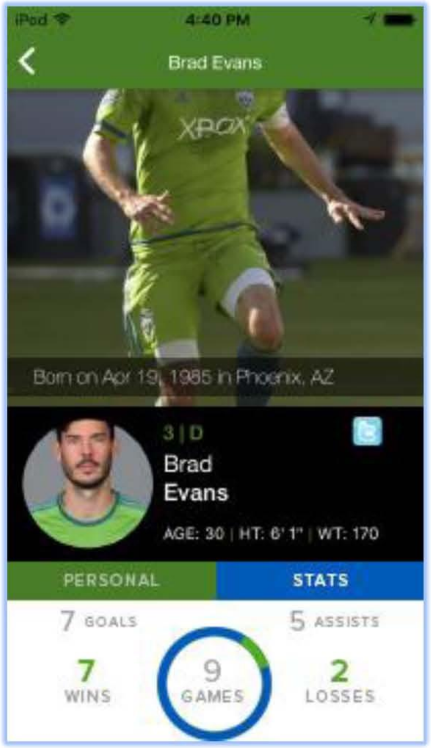

iPod
4:39 PM
Brad Evans
Born on Apr 19, 1985 in Phoenix, AZ
3 | D
Brad
Evans
AGE: 30 | HT: 6' 1" | WT: 170


iPod
4:40 PM
Brad Evans
Born on Apr 19, 1985 in Phoenix, AZ
3 | D
Brad
Evans
AGE: 30 | HT: 6' 1" | WT: 170
Career Statistics
24
1
Games 196
Starts 171
Minutes 15005
Goals 23
Assists 28
Shots 197
Shots on Goal 69
Fouls C/S 148 / 146
Corner Kicks 41
Offsides 20



Sounders Treatment Player Card

Here are screenshots of the player card with the tab/toggle. These screenshots also show other changes (graphs for the stats for example), but the "Personal" and "Stats" are the tabs to implement.


iPod
4:40 PM
Brad Evans
Born on Apr 19, 1985 in Phoenix, AZ
3 | D
Brad
Evans
AGE: 30 | HT: 6' 1" | WT: 170
PERSONAL
STATS
7 GOALS
5 ASSISTS
7
WINS
9
GAMES
2
LOSSES

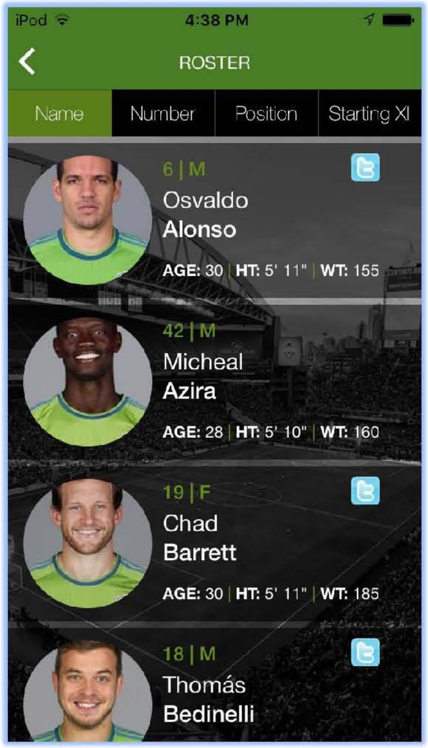

iPod
4:41 PM
Brad Evans
3 D-M
Brad
Evans
PERSONAL
STATS
18 GOALS
26 ASSISTS
145
WINS
167
GAMES
42
LOSSES
159 SHOTS
50 SHOTS ON GOAL
9777
STARTING
13230
MINUTES
0
SUBSTITUTE
21
0

2) Bigger Photos, Scroll left and right with Roster on Bottom of page

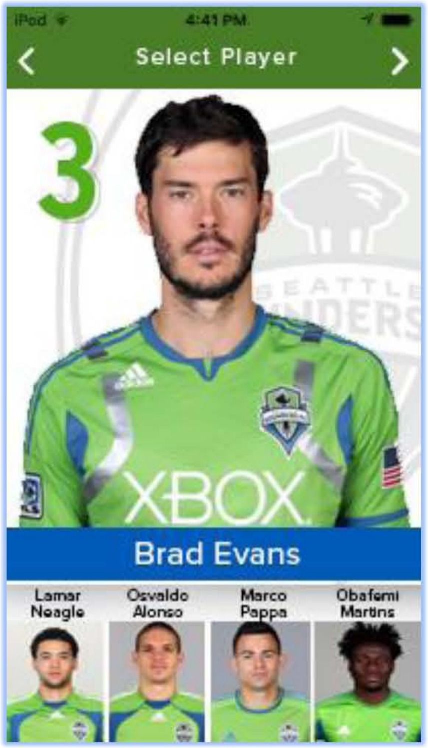
Sounders Current Roster List Page

The current implementation shows the players listed vertically with about 4 players shown at a time. The headshots are small, with a little bit of info listed for each player (age, height, weight).

Instead, **Real Madrid** uses a large picture of a default player with the navigation of the roster horizontally on the bottom. The Real Madrid app uses very little text on describing the player. But the page pops with a large, clear photo of the player. Navigation to other players is easy to see immediately with the other smaller photos of other players on the roster at the bottom of the page.


iPod
4:38 PM
ROSTER
Name
Number
Position
Starting XI
6 | M
Osvaldo
Alonso
AGE: 30 | HT: 5' 11" | WT: 155
42 | M
Micheal
Azira
AGE: 28 | HT: 5' 10" | WT: 160
19 | F
Chad
Barrett
AGE: 30 | HT: 5' 11" | WT: 185
18 | M
Thomás
Bedinelli

Sounders Treatment Roster List Page

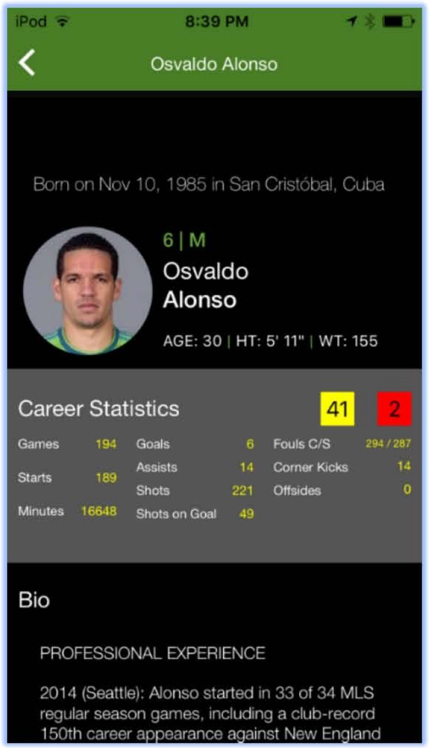


3) Graphs for Stats

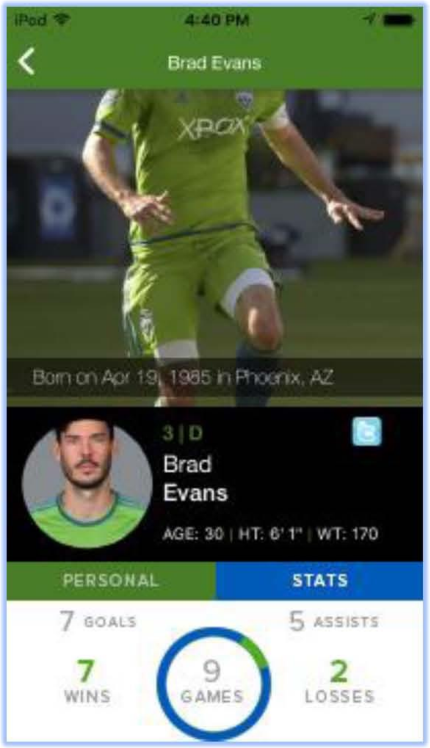
Current Sounders Implementation of Stats on Player Card

The current of the stats is static text on a great box on the player card. The number of stats given by the Sounders app is average compared to other soccer iOS apps, but others have a few graphics included which makes the page more attractive.

- **FC Barcelona** - Not an overwhelming amount of stats. Only 5 basic categories. And only using current season stats. Uses circle graph for wins/losses and starting/substitute.
- **One Football** – Only one graph here on the stats here (games played/games), but the information is layed out so that they are more visually appealing than the straight text box for the Sounders
- **Brooklyn Nets** - The charts and graphs for the Career has the statistics shown per year, whereas the same graphs in the current season use other teams as the pivot. Charts of stats are scrollable left and right as well allowing to see more columns. This isn't obvious in the UI so it is possible many users don't discover. This is a site with a ton of stats similar to an MLB site, or aggregator (ESPN for example)
- **Real Madrid** - The formatting/presentation of the stats is really cool for Real Madrid. There's a filter for looking at the stats in different competitions. There's an option to look at the stats as a trend as well as a ranking among other teammates.



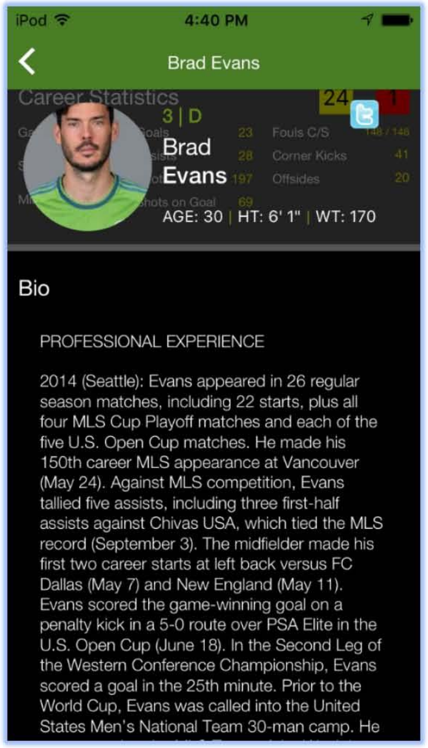
Sounders Treatment Stats on Player Card



4) Link to Jersey from Player Card

Current Sounders implementation of Player Card

The current player card does not provide a direct link to the store to buy a jersey. This is something that should drive revenue for the sounders, and is helpful to users since it provides a quick way to get to the store for a jersey without having to find the link to the team store and searching.


iPod
4:40 PM
Brad Evans
Career Statistics
24
1
3 | D
Brad
Evans
Fouls C/S
Corner Kicks 41
Offsides 20
AGE: 30 | HT: 6' 1" | WT: 170
Bio
PROFESSIONAL EXPERIENCE
2014 (Seattle): Evans appeared in 26 regular season matches, including 22 starts, plus all four MLS Cup Playoff matches and each of the five U.S. Open Cup matches. He made his 150th career MLS appearance at Vancouver (May 24). Against MLS competition, Evans tallied five assists, including three first-half assists against Chivas USA, which tied the MLS record (September 3). The midfielder made his first two career starts at left back versus FC Dallas (May 7) and New England (May 11). Evans scored the game-winning goal on a penalty kick in a 5-0 route over PSA Elite in the U.S. Open Cup (June 18). In the Second Leg of the Western Conference Championship, Evans scored a goal in the 25th minute. Prior to the World Cup, Evans was called into the United States Men's National Team 30-man camp. He

Current Sounders Implementation of Stats on Player Card

Appendix H: **TEDSRate Survey – Sounders FC mobile App for Player Information Scenario**

Sports Mobile App Comparison - Non-Expert - Player Information

Welcome

This assessment is a component of the Purposeful Sampling Research Project and the University of Washington Information School. The work group's aim is to refine a methodology to measure the usability of content and information artifacts. It has been used to evaluate mobile applications with a concentration on professional sport team mobile applications, but it is now being adapted to evaluate emergency management information systems, specifically WebEOC. The methodology is based on the Taylor-Eisenberg-Dirks-Scholl (TEDS) information artifact value factorization framework.

Although these assessments are short, you do not need the do them all at once. Your changes are saved so you can take as much time as you need!

The focus of this assessment is to analyze the Sounders FC Mobile App

Thank you in advance for taking the time to take part in this evaluation!

The TEDS Purposeful Sampling Research Group

Start

Sports Mobile App Comparison - Non-Expert - Player Information

Hello New User

Please complete your registration or, if you have already registered with us before, sign in

Please provide us with an email and password so that we may track you across individual ratings.

An email is required

Email

Password

Confirm

Back Sign Up

Sports Mobile App Comparison - Non-Expert - Player Information

Demographic Questions

Privacy Policy

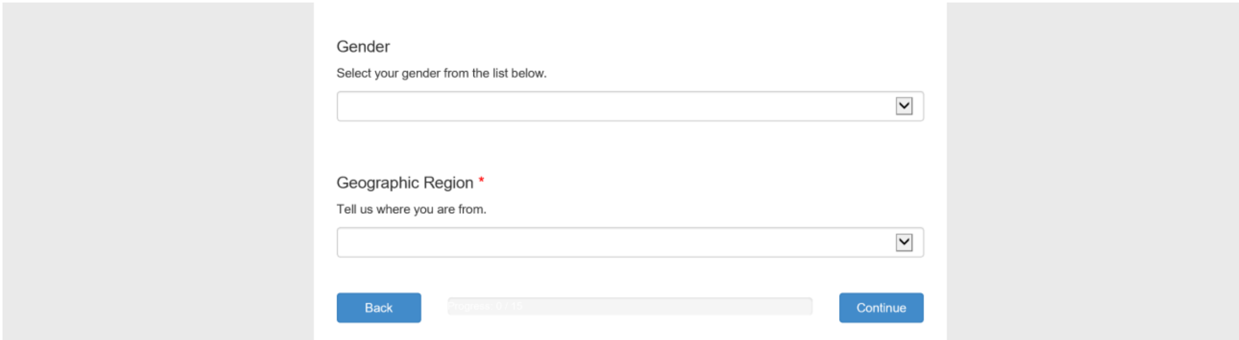
Your personal information is used to better understand the needs and preferences of groups of individuals with similar characteristics. We do not sell, trade, or otherwise transfer to outside parties your personally identifiable information. After the research project is completed the information you provided will be completely deleted.

Education

What is the highest level of school you have completed?

Age *

Select your age range from the list below.



Gender

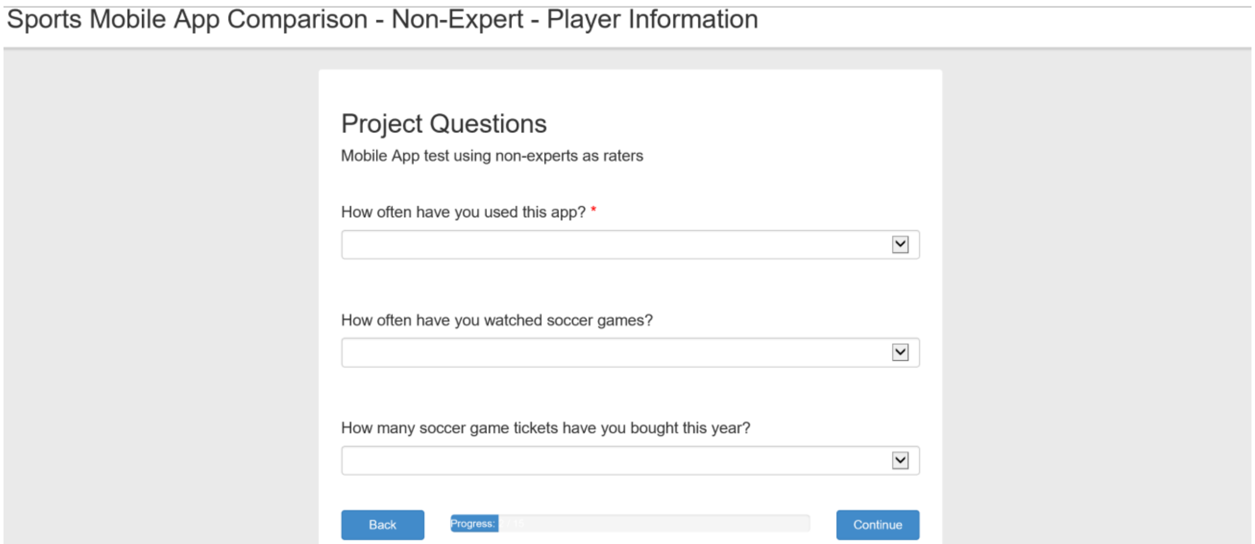
Select your gender from the list below.

Geographic Region *

Tell us where you are from.

Back

Continue



Sports Mobile App Comparison - Non-Expert - Player Information

Project Questions

Mobile App test using non-experts as raters

How often have you used this app? *

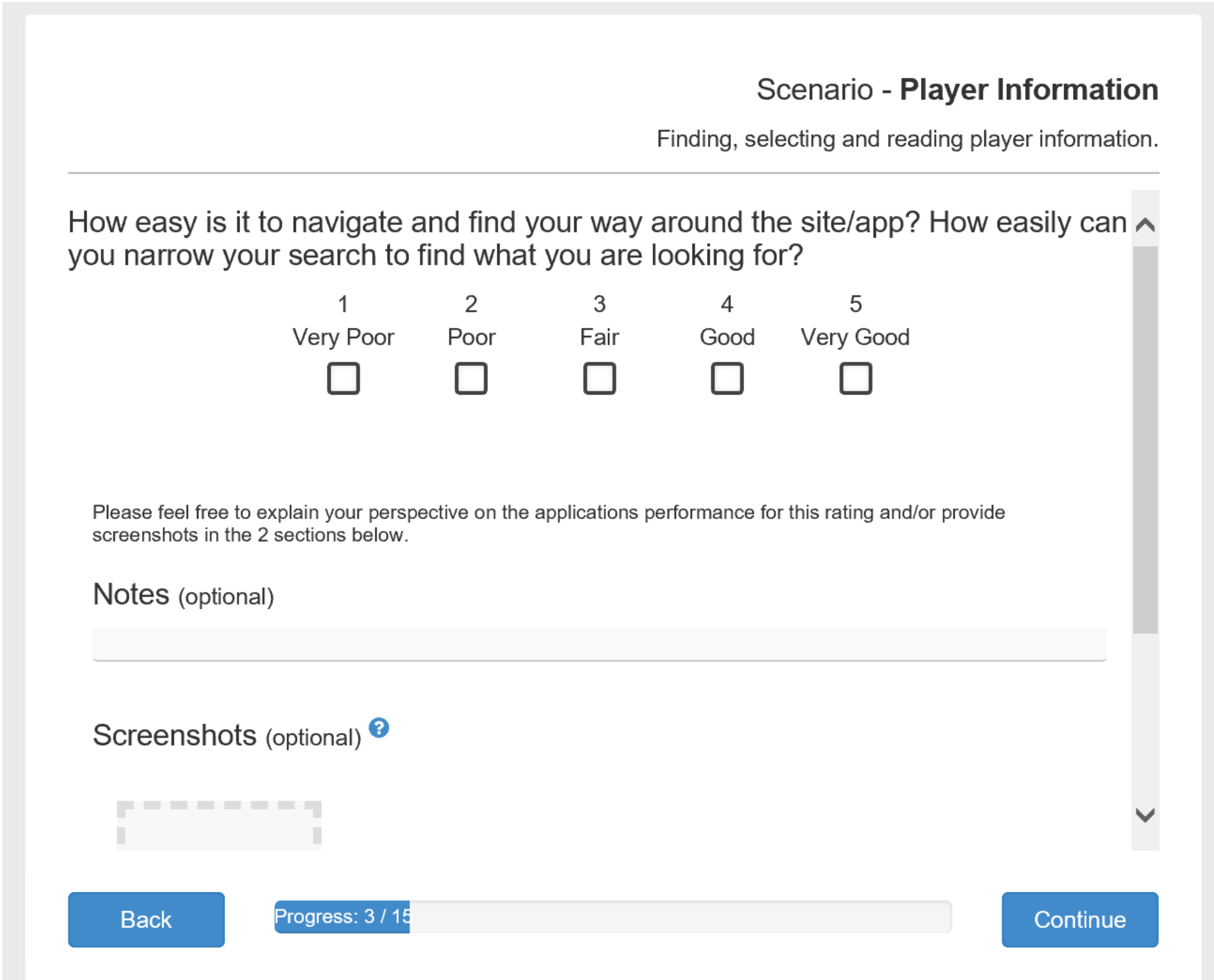
How often have you watched soccer games?

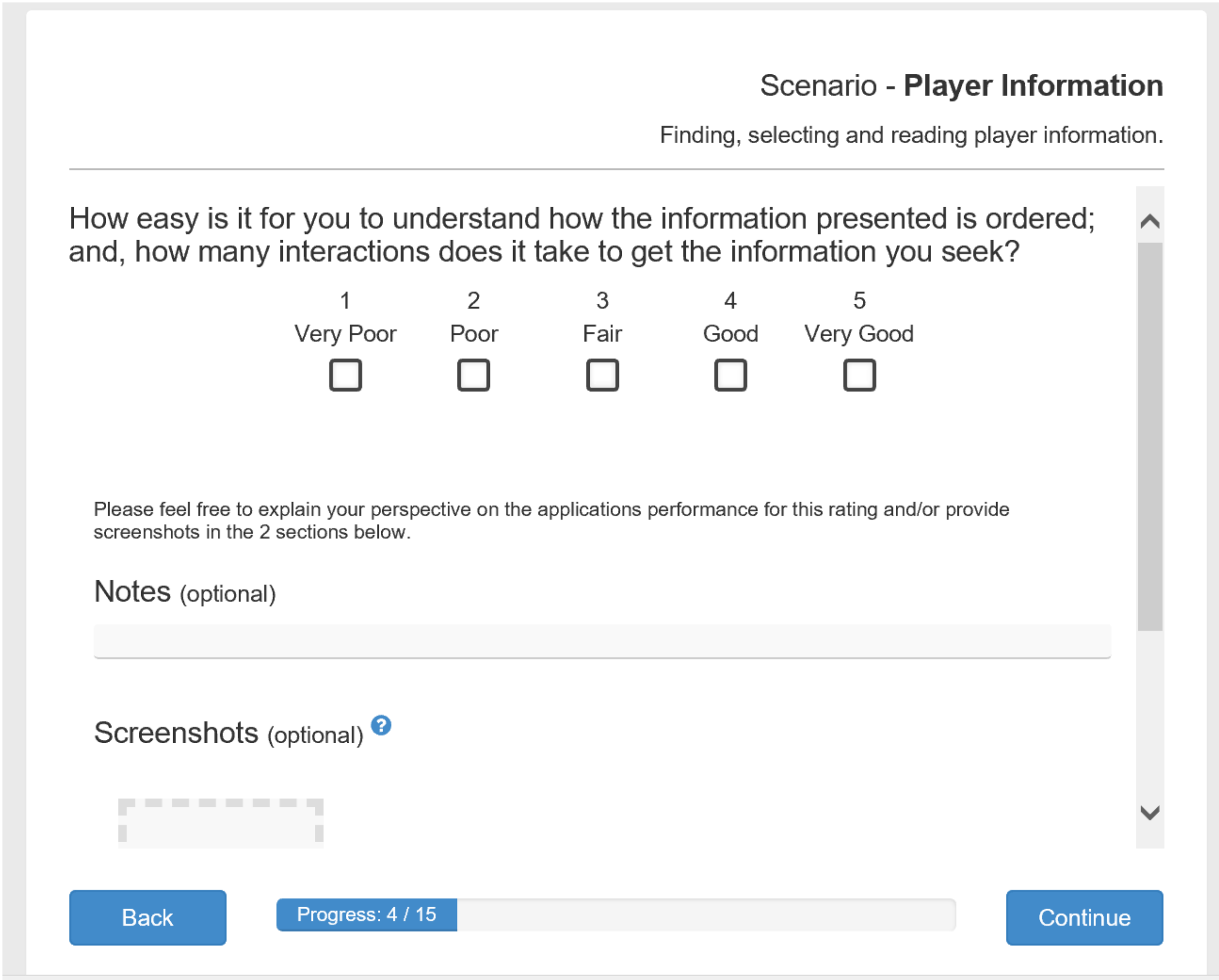
How many soccer game tickets have you bought this year?

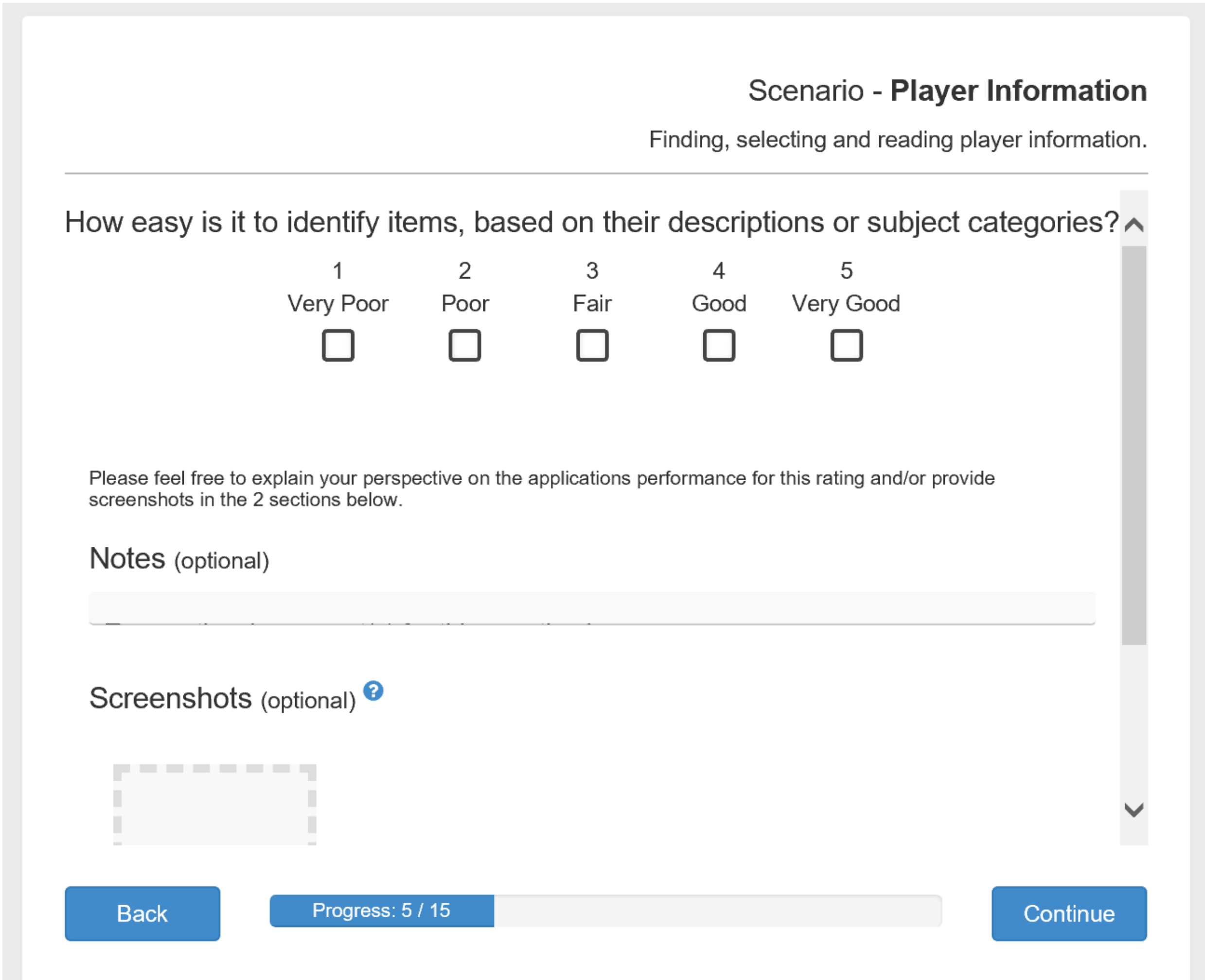
Back

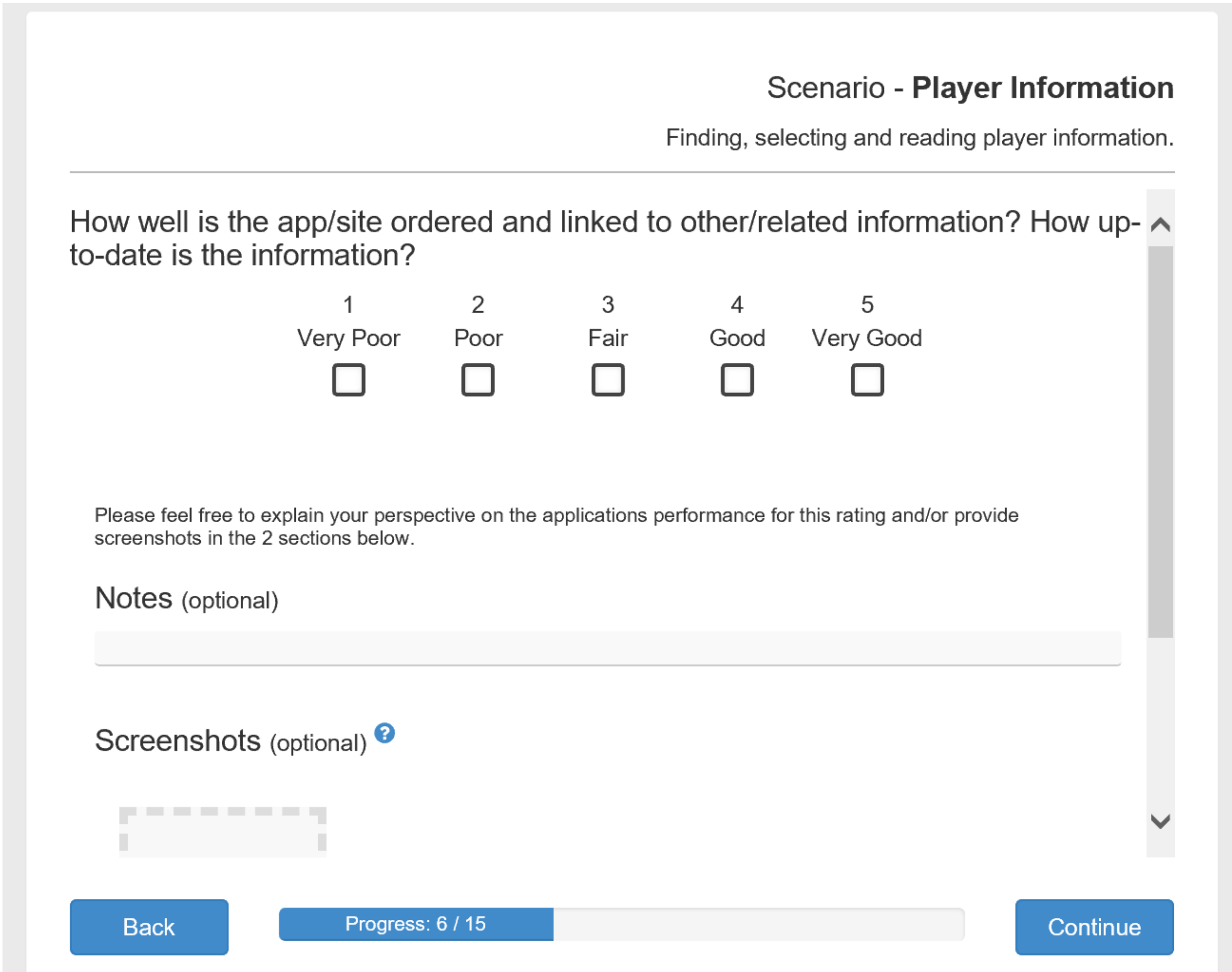
Progress:

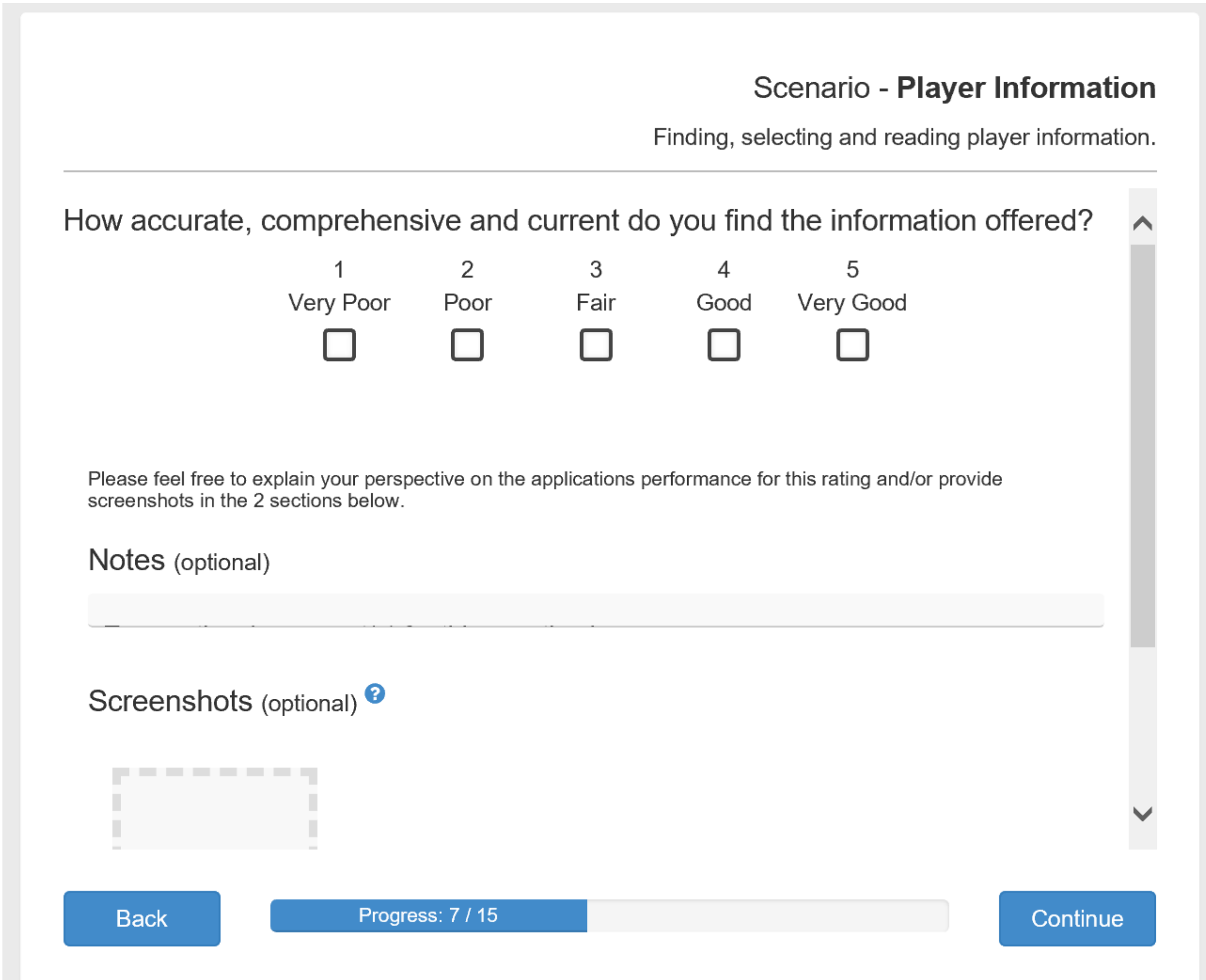
Continue

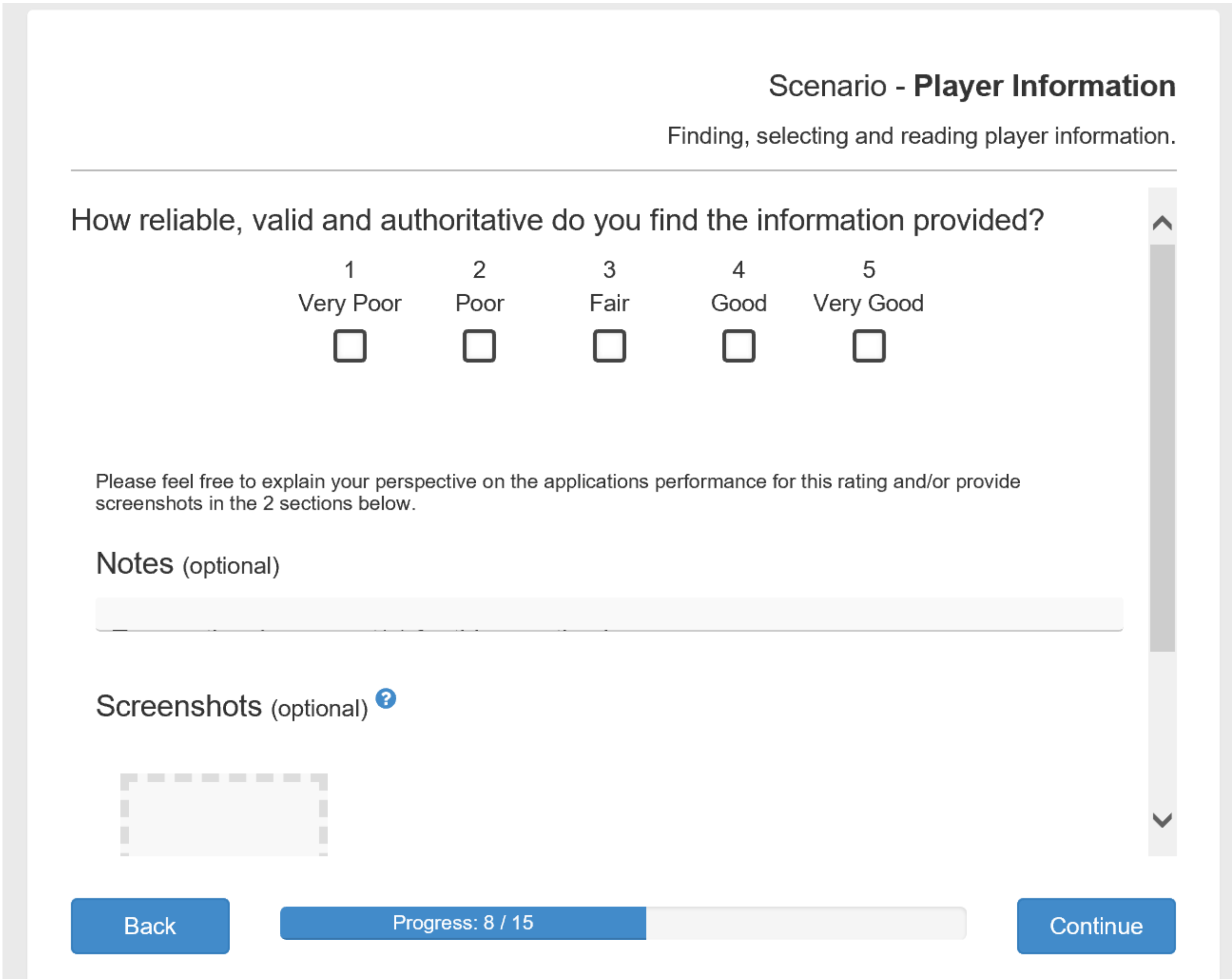

Scenario - **Player Information**
Finding, selecting and reading player information.
How easy is it to navigate and find your way around the site/app? How easily can you narrow your search to find what you are looking for?
1 Very Poor
2 Poor
3 Fair
4 Good
5 Very Good
Please feel free to explain your perspective on the applications performance for this rating and/or provide screenshots in the 2 sections below.
Notes (optional)
Screenshots (optional)
Back
Progress: 3 / 15
Continue


Scenario - Player Information
Finding, selecting and reading player information.
How easy is it for you to understand how the information presented is ordered; and, how many interactions does it take to get the information you seek?
1 Very Poor
2 Poor
3 Fair
4 Good
5 Very Good
Please feel free to explain your perspective on the applications performance for this rating and/or provide screenshots in the 2 sections below.
Notes (optional)
Screenshots (optional)
Back
Progress: 4 / 15
Continue

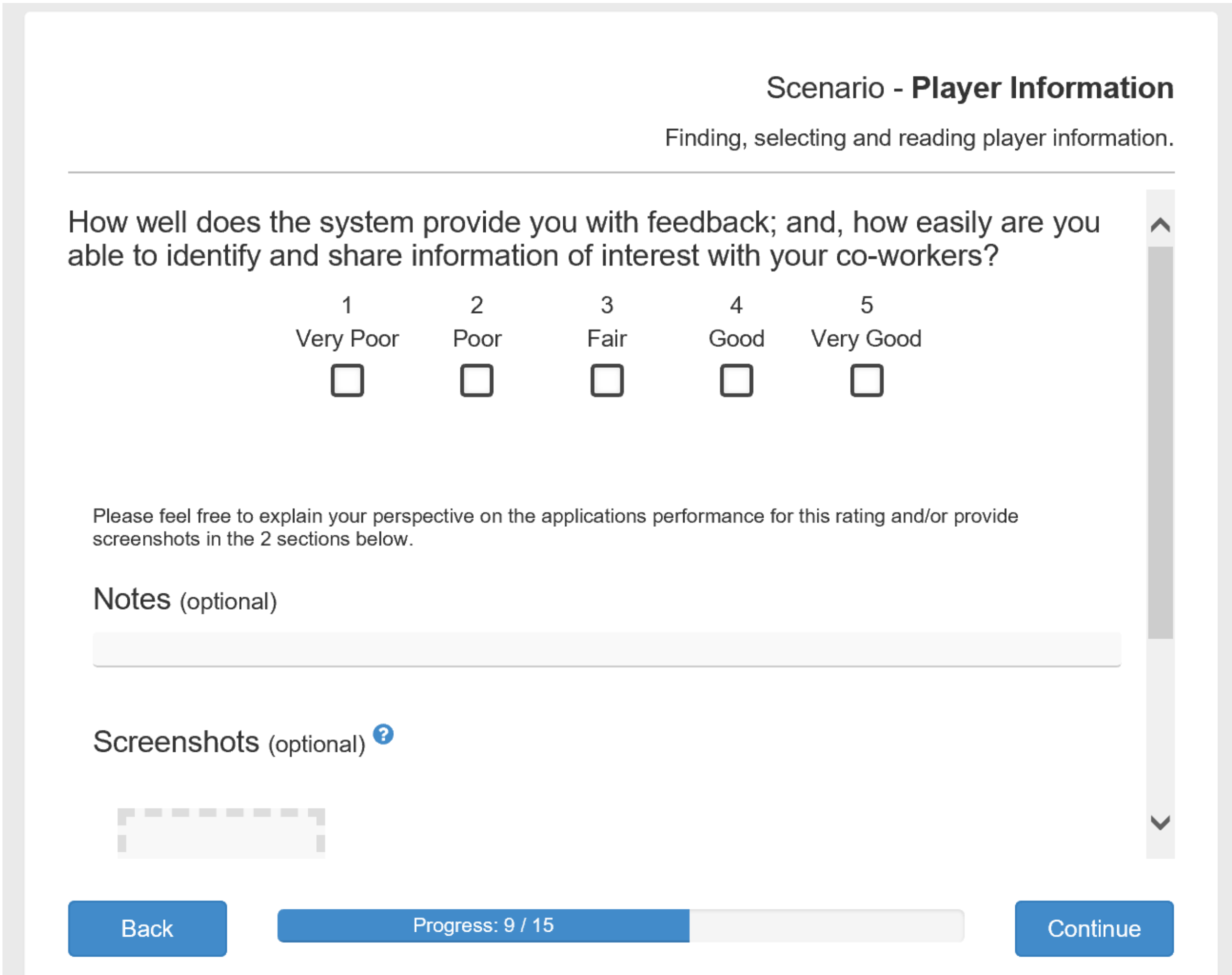

Scenario - Player Information
Finding, selecting and reading player information.
How easy is it to identify items, based on their descriptions or subject categories?
1 Very Poor
2 Poor
3 Fair
4 Good
5 Very Good
Please feel free to explain your perspective on the applications performance for this rating and/or provide screenshots in the 2 sections below.
Notes (optional)
Screenshots (optional)
Back
Progress: 5 / 15
Continue


Scenario - Player Information
Finding, selecting and reading player information.
How well is the app/site ordered and linked to other/related information? How up-to-date is the information?
1 Very Poor
2 Poor
3 Fair
4 Good
5 Very Good
Please feel free to explain your perspective on the applications performance for this rating and/or provide screenshots in the 2 sections below.
Notes (optional)
Screenshots (optional)
Back
Progress: 6 / 15
Continue


Scenario - **Player Information**
Finding, selecting and reading player information.
How accurate, comprehensive and current do you find the information offered?
1 Very Poor
2 Poor
3 Fair
4 Good
5 Very Good
Please feel free to explain your perspective on the applications performance for this rating and/or provide screenshots in the 2 sections below.
Notes (optional)
Screenshots (optional)
Back
Progress: 7 / 15
Continue



Scenario - Player Information
Finding, selecting and reading player information.
How reliable, valid and authoritative do you find the information provided?
1 Very Poor
2 Poor
3 Fair
4 Good
5 Very Good
Please feel free to explain your perspective on the applications performance for this rating and/or provide screenshots in the 2 sections below.
Notes (optional)
Screenshots (optional)
Back
Progress: 8 / 15
Continue



Scenario - **Player Information**

Finding, selecting and reading player information.

How well does the system provide you with feedback; and, how easily are you able to identify and share information of interest with your co-workers?

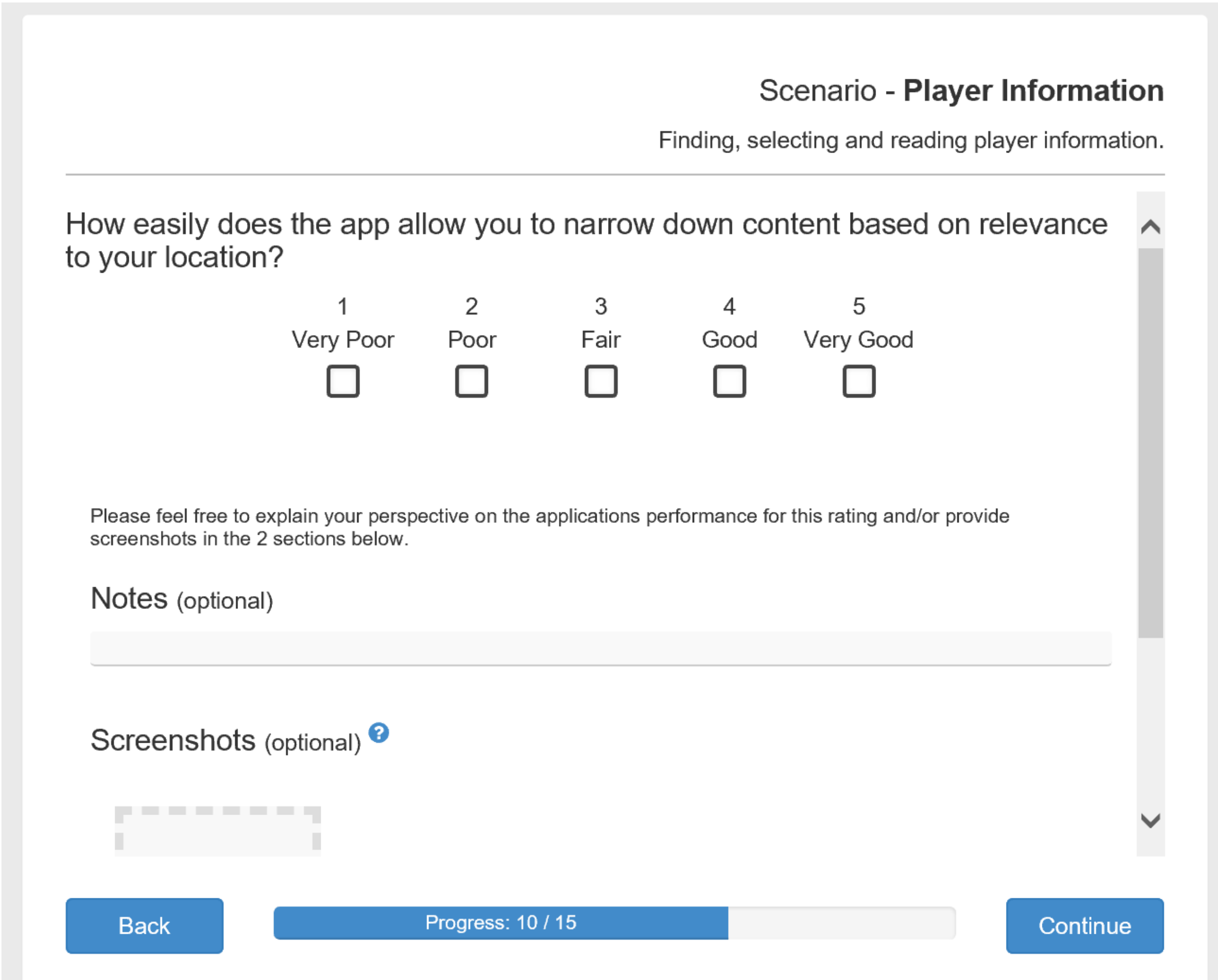
1	2	3	4	5
Very Poor	Poor	Fair	Good	Very Good
☐	☐	☐	☐	☐

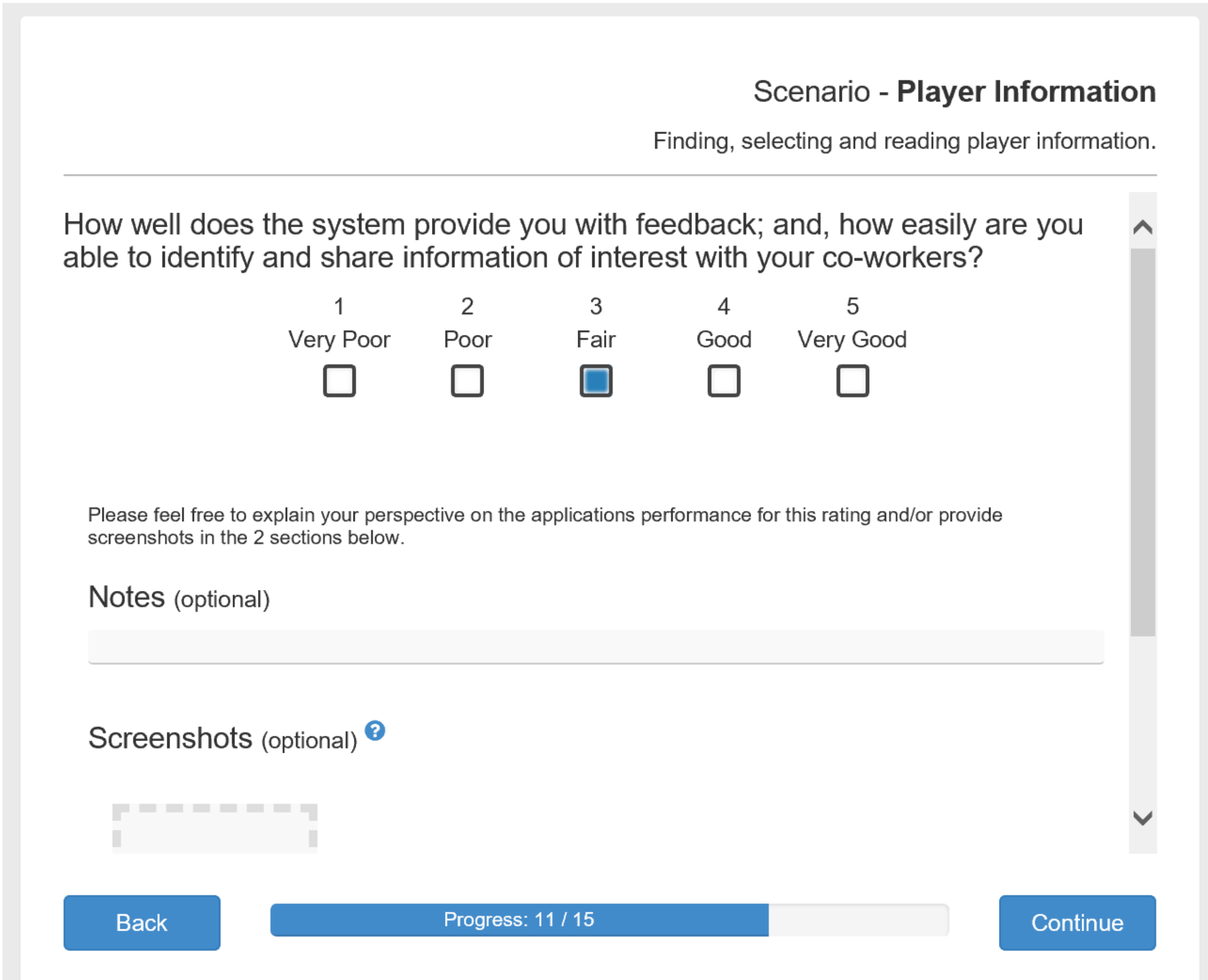
Please feel free to explain your perspective on the applications performance for this rating and/or provide screenshots in the 2 sections below.

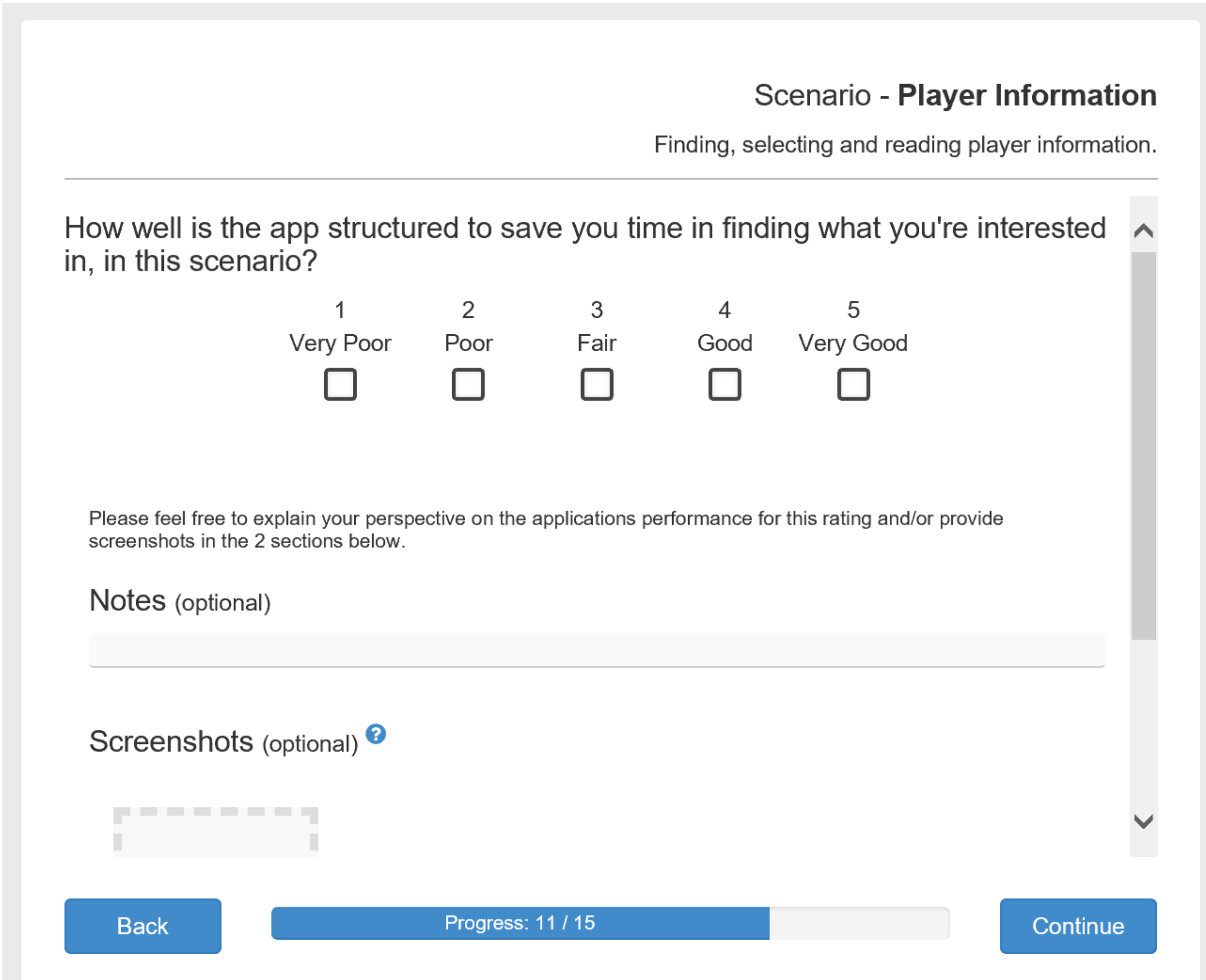
Notes (optional)

Screenshots (optional)

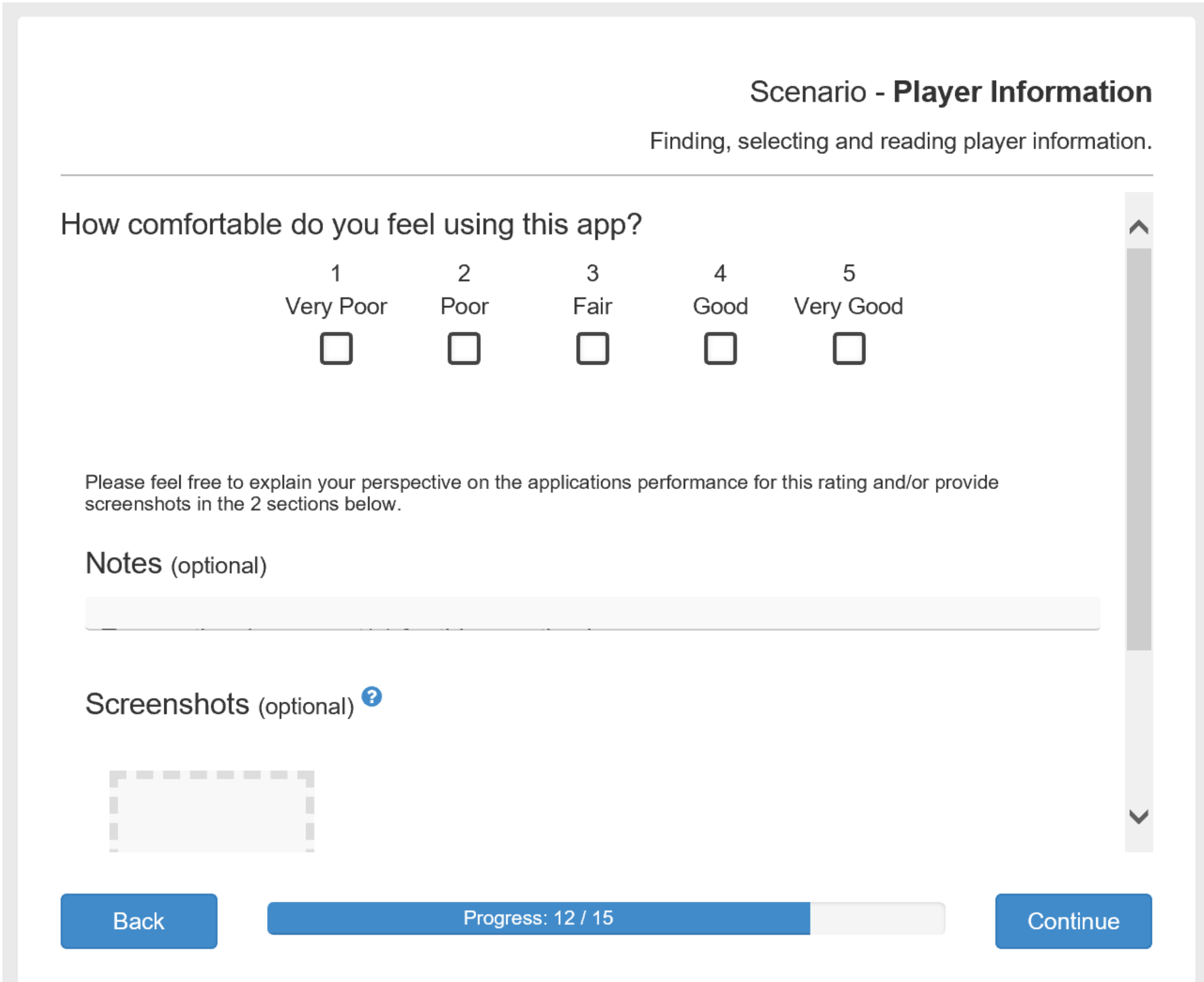
Back | Progress: 9 / 15 | Continue

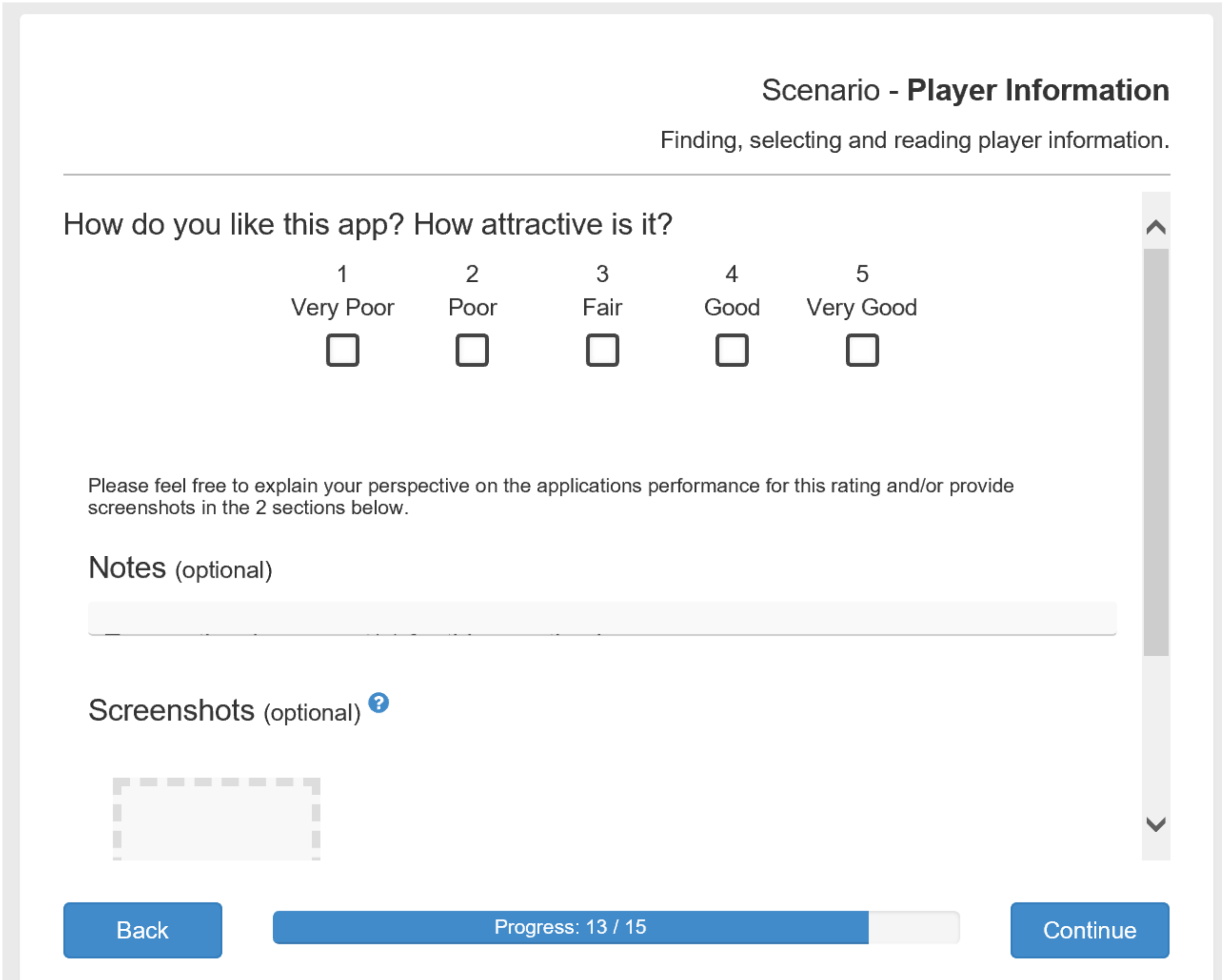

Scenario - Player Information
Finding, selecting and reading player information.
How easily does the app allow you to narrow down content based on relevance to your location?
1
Very Poor
2
Poor
3
Fair
4
Good
5
Very Good
Please feel free to explain your perspective on the applications performance for this rating and/or provide screenshots in the 2 sections below.
Notes (optional)
Screenshots (optional)
Back
Progress: 10 / 15
Continue

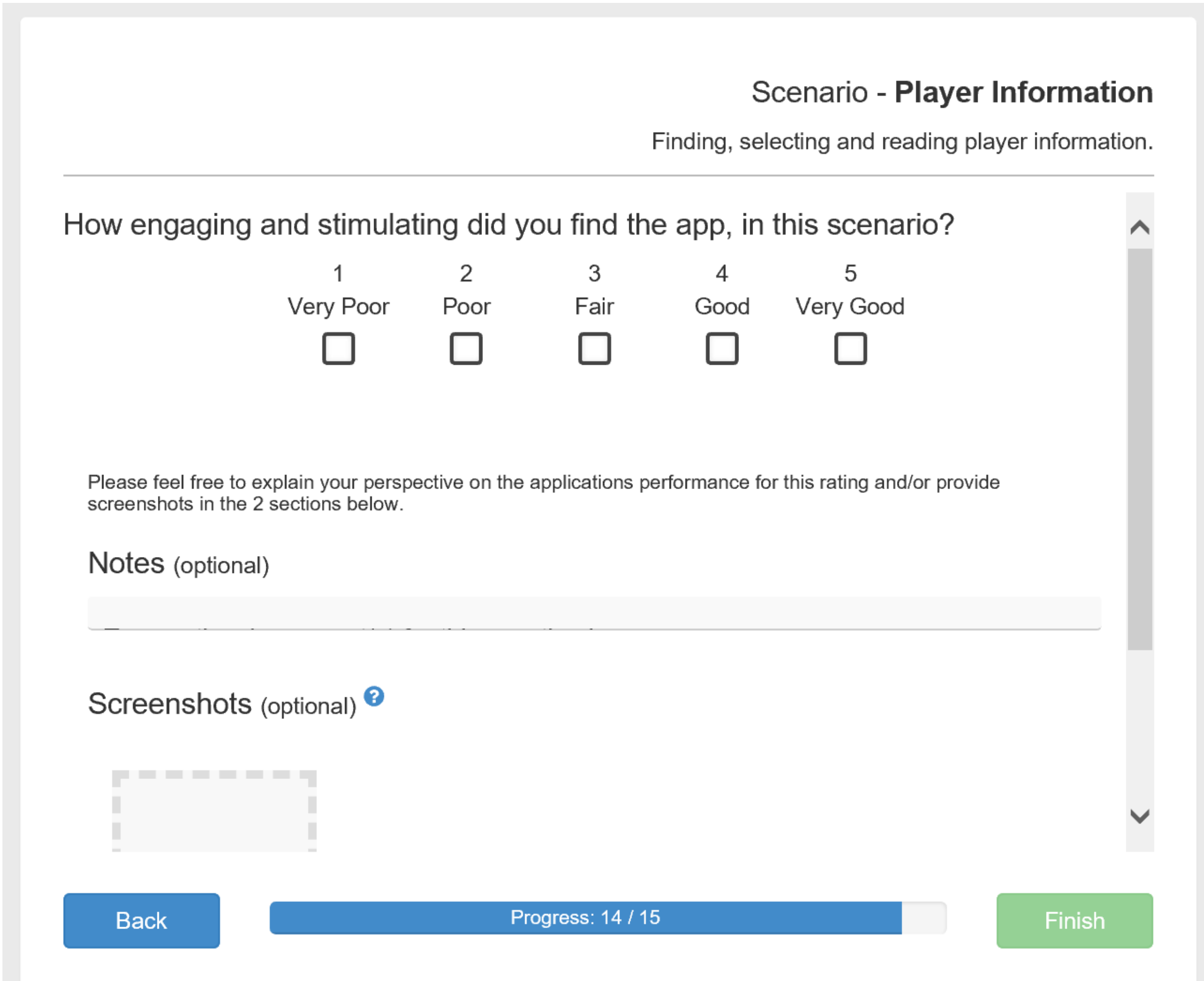

Scenario - **Player Information**
Finding, selecting and reading player information.
How well does the system provide you with feedback; and, how easily are you able to identify and share information of interest with your co-workers?
1 Very Poor
2 Poor
3 Fair
4 Good
5 Very Good
Please feel free to explain your perspective on the applications performance for this rating and/or provide screenshots in the 2 sections below.
Notes (optional)
Screenshots (optional)
Back
Progress: 11 / 15
Continue

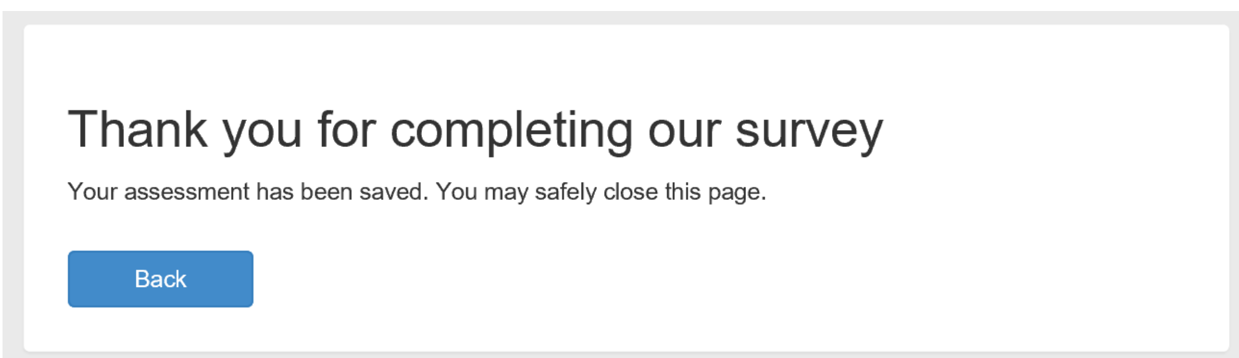


Scenario - **Player Information**
Finding, selecting and reading player information.
How well is the app structured to save you time in finding what you're interested in, in this scenario?
1 Very Poor
2 Poor
3 Fair
4 Good
5 Very Good
Please feel free to explain your perspective on the applications performance for this rating and/or provide screenshots in the 2 sections below.
Notes (optional)
Screenshots (optional)
Back
Progress: 11 / 15
Continue


Scenario - **Player Information**
Finding, selecting and reading player information.
How comfortable do you feel using this app?
1 Very Poor
2 Poor
3 Fair
4 Good
5 Very Good
Please feel free to explain your perspective on the applications performance for this rating and/or provide screenshots in the 2 sections below.
Notes (optional)
Screenshots (optional)
Back
Progress: 12 / 15
Continue


Scenario - Player Information
Finding, selecting and reading player information.
How do you like this app? How attractive is it?
1 Very Poor
2 Poor
3 Fair
4 Good
5 Very Good
Please feel free to explain your perspective on the applications performance for this rating and/or provide screenshots in the 2 sections below.
Notes (optional)
Screenshots (optional)
Back
Progress: 13 / 15
Continue


Scenario - Player Information
Finding, selecting and reading player information.
How engaging and stimulating did you find the app, in this scenario?
1 Very Poor
2 Poor
3 Fair
4 Good
5 Very Good
Please feel free to explain your perspective on the applications performance for this rating and/or provide screenshots in the 2 sections below.
Notes (optional)
Screenshots (optional)
Back
Progress: 14 / 15
Finish


Thank you for completing our survey
Your assessment has been saved. You may safely close this page.
Back

Appendix J: Phase 1 Expert Ratings of Sounders FC iOS Application for Player Information and for Schedule, Results, and League

Player Info		**Phase 1**						
Criteria	**Values**	*Rater 1*	*Rater 2*	*Rater 3*	*Rater 4*	*Rater 5*	*Std Dev*	*Average*
Ease of Use	**Navigation and Findability** Browsing/searchability, Mediation, Orientation, Simplicity	4.00	4.00	**3.50**	**4.50**	4.00	0.35	4.00
Ease of Use	**Structure** Formatting/Presentation, Order, Accessibility	3.50	3.50	**3.00**	**4.00**	**4.00**	0.42	3.60
Noise Reduction	**Identity** Item Identification, Subject description, Summary, Selectivity, Precision	3.50	**3.00**	3.50	**4.00**	3.50	0.35	3.50
Noise Reduction	**Parsimony** Linkage/Referral, Order, Novelty	3.50	3.50	**3.00**	**4.50**	3.50	0.55	3.60
Quality	**Completeness** Accuracy, Comprehensiveness, Currency	3.00	3.00	**3.50**	3.00	**3.50**	0.27	3.20
Quality	**Trustworthiness** Reliability, Validity, Authority	3.50	3.50	**4.00**	**4.00**	3.50	0.27	3.70
Adaptability	**Interaction** Contextuality, Transaction, Feedback, Community/Social networking	3.00	3.00	3.00	**4.00**	3.50	0.45	3.30
Adaptability	**Customization** Flexibility, Simplicity, Trust, Individualization, Localization, Privacy	**3.00**	2.50	2.50	**2.00**	**3.00**	0.42	2.60
Performance	**Savings** Cost savings, Time savings	**3.00**	**3.00**	**3.00**	**3.00**	**3.00**	0.00	3.00
Performance	**Confidence** Security, Safety	3.50	3.00	3.00	**4.00**	3.00	0.45	3.30
Affection	**Attractiveness** Aesthetics, Satisfaction/rewarding/ incenting	3.50	3.00	**2.50**	3.00	**4.00**	0.57	3.20
Affection	**Enjoyment** Entertainment, Engagement, Stimulation	3.50	3.00	**2.50**	3.00	**4.00**	0.57	3.20
	Total averages	3.38	3.17	**3.08**	**3.58**	3.54	0.22	3.35

Schedule, Results and League		**Phase 1**						
Criteria	**Values**	***Rater 1***	***Rater 2***	***Rater 3***	***Rater 4***	***Rater 5***	***Std Dev***	***Average***
Ease of Use	**Navigation and Findability** Browsing/searchability, Mediation, Orientation, Simplicity	4.00	4.00	**3.50**	4.00	4.00	0.22	3.90
Ease of Use	**Structure** Formatting/Presentation, Order, Accessibility	**4.00**	3.50	3.50	**3.00**	**4.00**	0.42	3.60
Noise Reduction	**Identity** Item Identification, Subject description, Summary, Selectivity, Precision	4.00	**2.50**	3.50	4.00	4.00	0.65	3.60
Noise Reduction	**Parsimony** Linkage/Referral, Order, Novelty	3.50	3.00	3.00	3.00	**4.00**	0.45	3.30
Quality	**Completeness** Accuracy, Comprehensiveness, Currency	**4.00**	3.00	**2.50**	3.00	3.00	0.55	3.10
Quality	**Trustworthiness** Reliability, Validity, Authority	4.00	**3.00**	4.00	4.00	**3.00**	0.55	3.60
Adaptability	**Interaction** Contextuality, Transaction, Feedback, Community/Social networking	2.50	2.00	2.00	2.00	**3.00**	0.45	2.30
Adaptability	**Customization** Flexibility, Simplicity, Trust, Individualization, Localization, Privacy	2.50	**2.00**	2.50	**2.00**	**3.00**	0.42	2.40
Performance	**Savings** Cost savings, Time savings	**4.00**	3.00	**4.00**	3.00	3.00	0.55	3.40
Performance	**Confidence** Security, Safety	**3.00**	**3.00**	**3.00**	**3.00**	**3.00**	0.00	3.00
Affection	**Attractiveness** Aesthetics, Satisfaction/rewarding/ incenting	**4.00**	3.00	3.00	3.00	**4.00**	0.55	3.40
Affection	**Enjoyment** Entertainment, Engagement, Stimulation	**4.00**	3.00	3.00	3.00	**4.00**	0.55	3.40
	Total averages	**3.63**	**2.92**	3.13	3.08	3.50	0.30	3.25

Appendix K: Phase 1 User Ratings of Sounders FC iOS Application

Player Information (Phase 1, part 1 of 3)

AssessmentID	Navigation and Findability	Structure	Identity	Substantiality	Completeness	Trustworthiness	Interaction	Customization	Savings	Confidence	Attractiveness	Enjoyment	How often have you used this app?	How often have you watched soccer games?	How many soccer game tickets have you bought this year?	Age	Gender	Geographic Region	Education
407	3	3	4	4	3	5	4	4	4	4	3	4	Almost every day	I try to watch every game	16+	40-49	Male	North America	Doctorate
405	5	5	5	5	5	5	5	5	4	5	5	5	Never	I try to watch every game	2-5	40-49	Male	North America	Bachelor's degree
408	3	4	4	4	5	5	5	4	5	5	5	5	Regularly	I try to watch every game	0	30-39	Female	Central America	High school
478	2	4	4	2	4	4	3	2	4	4	3	3	Occasionally	I try to watch every game	16+	50-59	Female	North America	Bachelor's degree
401	3	3	5	4	4	4	4	4	4	4	4	4	Regularly	I try to watch every game	16+	30-39	Male	North America	Bachelor's degree
328	3	3	4	4	3	4	2	4	3	3	4	4	Never	Regularly	2-5	20-29	Male	North America	Master's degree
333	4	4	5	3	4	5	4	3	4	3	4	5	Occasionally	Regularly	1	20-29	Male	South Asia	Bachelor's degree
391	3	4	4	4	4	4	4	4	4	5	4	4	Regularly	Regularly	16+	30-39	Male	North America	Some college
482	3	2	4	4	4	3	3	3	2	3	2	3	Regularly	I try to watch every game	16+	20-29		North America	Some college
411	4	4	4	3	5	4	4	4	4	4	5	5	Regularly	I try to watch every game	2-5	20-29	Female	North America	Bachelor's degree
410	4	4	5	4	3	5	3	3	5	5	4	4	Regularly	Regularly	10-15	40-49	Female	North America	Some college
472	5	5	5	5	5	5	4	3	4	5	5	5	Regularly	I try to watch every game	0	30-39	Male	North America	High school
444	1	1	5	1	3	1	1	3	5	1	1	1	Never	I try to watch every game	16+	40-49	Male	North America	Some college
442	3	4	4	4	4	4	3	4	4	5	4	4	Never	Occasionally	1	20-29	Female	North America	Bachelor's degree
439	4	4	5	4	5	4	3	4	4	5	4	4	Occasionally	Regularly	1	50-59	Male	North America	Some college
385	4	4	4	4	4	5	2		2	5	4	3	Regularly	I try to watch every game	16+	50-59	Female	North America	Master's degree
449	2	3	4	4	4	4	3	4	4	4	3	2	Rarely	Regularly	2-5	Under 20	Female	North America	Bachelor's degree
423	3	4	4	3	3	3	2	2	3	3	4	3	Regularly	I try to watch every game	2-5	20-29	Female	North America	Bachelor's degree
459	3	3	4	3	5	4	4	4	4	3	4	3	Occasionally	Regularly	16+	50-59	Male	North America	Doctorate
451	3	4	4	3	4	3	4	4	4	4	4	4	Rarely	Regularly	1	Under 20	Female	North America	High school
431	3	3	5	4	4	5	4	5	4	4	4	4	Regularly	Regularly	2-5	30-39	Male	North America	Some college
492	4	4	4	4	4	5	4	4	4	4	4	3	Regularly	I try to watch every game	2-5	30-39	Female	North America	Some college
494	4	4	4	4	4	4	4	4	4	4	4	4	Regularly	Regularly	2-5	30-39	Female	North America	Some college

(Continued on next page)

(Continued from prior page)

Player Information (Phase 1, part 2 of 3)

AssessmentID	Navigation and Findability	Structure	Identity	Substantiality	Completeness	Trustworthiness	Interaction	Customization	Savings	Confidence	Attractiveness	Enjoyment	How often have you used this app?	How often have you watched soccer games?	How many soccer game tickets have you bought this year?	Age	Gender	Geographic Region	Education
504	3	4	3	4	4	4	4	4	3	4	4	4	Occasionally	Regularly	2-5	30-39	Male	North America	Some college
502	3	3	3	4	4		4	3	3				Never	I try to watch every game	2-5	40-49	Female	North America	Bachelor's degree
506	5	4	3	4	4	4	3	3	5	5	4	3	Occasionally	I try to watch every game	16+	20-29	Female	North America	Some college
510	4	4	5	5	4	5	3	4	5	5	5	3	Rarely	I try to watch every game	1	30-39	Male	North America	Master's degree
521	4	4	3	5	4	4	4	4	3	4	4	3	Never	Occasionally	0	20-29	Female	North America	Bachelor's degree
516	4	5	5	4	3	4	4	4	4	4	4	4	Regularly	I try to watch every game	16+	40-49	Female	North America	Bachelor's degree
519	5	5	5	5		5	5	5	5	4	5	5	Occasionally	I try to watch every game	2-5	20-29	Male	North Africa	Some college
548	2	1	2	3	3	3	2	3	1	3	4	3	Regularly	I try to watch every game	16+	50-59		North America	Bachelor's degree
529	3	3	4	3	3	3	2	3	2	3	2	2	Occasionally	Occasionally	1	30-39	Female	North America	Bachelor's degree
530	5	5	4	5	5	5	5	5	5	5	5	4	Regularly	I try to watch every game	2-5	50-59	Female	North America	High school
546	5	4	5	5	5	5	5	5	5	5	5	5	Regularly	Regularly	10-15	Under 20	Male	North America	Some college
537	4	4	4	5	4	5	4	4	4	4	4	4	Occasionally	Regularly	6-9	30-39	Female	North America	Bachelor's degree
542	4	5	4	5	5	5	3	3	3	4	4	3	Regularly	I try to watch every game	16+	30-39	Female	North America	Master's degree
553	5	4	5	5	5	4	4	4	5	5	4	4	Almost every day	I try to watch every game	10-15	40-49	Female	North America	Bachelor's degree
562	4	4	4	3	4	4	3	2					Occasionally	I try to watch every game	10-15	20-29	Male	North America	Bachelor's degree
563	5	4	5	5	4	4	5	5	5	4	4	5	Occasionally	Regularly	6-9	30-39	Female	North America	Some college
556	4	3	2	4	5	4	3	2	3	3	3	4	Occasionally	Regularly	1	20-29	Female	North America	Some college
564	2	2	2	2	2	2	2	2	2	3	2	2	Regularly	I try to watch every game	16+	40-49	Male	North America	Bachelor's degree
596	3	3	4	2	3	4	4	4	4	4	4	4	Regularly	Regularly	1	50-59	Female	North America	Bachelor's degree
583	3	4	3	2	5	4	3	3	3	3	3	3	Occasionally	Regularly	16+	40-49	Male	North America	Bachelor's degree
586	2	2	3	3	2	3	2	2	1	1	1	1	Rarely	I try to watch every game	16+	30-39	Female	North America	Bachelor's degree

(Continued on next page)

(Continued from prior page)

Player Information (Phase 1, part 3 of 3)

AssessmentID	Navigation and Findability	Structure	Identity	Substantiality	Completeness	Trustworthiness	Interaction	Customization	Savings	Confidence	Attractiveness	Enjoyment	How often have you used this app?	How often have you watched soccer games?	How many soccer game tickets have you bought this year?	Age	Gender	Geographic Region	Education
589	4	4	4	3	4	5							Occasionally	Regularly	0	30-39		North America	Bachelor's degree
591	3	3	4	2	3	3	2	2	2	2	1	2	Rarely	Regularly	16+	50-59	Female	North America	Bachelor's degree
599	3	4	4	4	4	5	3	3	4	4	4	3	Rarely	Regularly	2-5	20-29	Female	North America	Some college
612	1	1	1	1	1	1	1	1	1	1	1	1	Occasionally	I try to watch every game	16+	30-39	Female	North America	Master's degree
607	2	2	3	3	3	3	2	3	3	3	2	2	Occasionally	Occasionally	0	40-49	Female	North America	Master's degree
613	4	4		5	5	4							Occasionally	I try to watch every game	16+	40-49		North America	Master's degree
616	4	4	4	5	5	5	4	4	5	4	3	3	Rarely	Regularly	2-5	20-29	Male	North America	Bachelor's degree
630	4	4	4	3	4	5	3	4	4	5	4	3	Never	Occasionally	2-5	20-29	Male	North America	Some college
631	5	5	4	5	4	5	3	5	5	5	5	4	Occasionally	Regularly	2-5	20-29	Female	North America	Some college
642	4	5	5	2	4	5							Occasionally	Regularly	6-9	Under 20			High school
637	2	2	3	2	3	3	2	2	1	1	1	2	Occasionally	I try to watch every game	16+	30-39	Male	North America	Some college
639	3	3	4	4	4	3	4	3	4	4	4	4	Regularly	I try to watch every game	0	Under 20	Female	North America	High school

Schedule Results League (Phase 1, part 1 of 2): (Continued on next page)

Navigation and Findability	Structure	Identity	Substantiality	Completeness	Trustworthiness	Interaction	Customization	Savings	Confidence	Attractiveness	Enjoyment	How often have you used this app?	How often have you watched soccer games?	How many soccer game tickets have you bought this year?	Age	Gender	Geographic Region	Education
4	4	4	3	4	4	3	3	3	3	4	4	Regularly	I try to watch every game	16+	30-39	Male	North America	Bachelor's degree
5	5	5	5	5	5	5	5	5	5	5	4	Never	I try to watch every game	2-5	40-49	Male	North America	Bachelor's degree
2	2	2	3	3	5	1	2	2	3	2	3	Regularly	I try to watch every game	16+	20-29	Male	North America	Some college
4	4	4	4	5	4	4	4	4	4	5	5	Regularly	I try to watch every game	2-5	20-29	Female	North America	Bachelor's degree
4	4	4	4	4	4	4	4	4	4	4	4	Regularly	Regularly	16+	30-39	Male	North America	Some college
4	3	4	5	4	3	4	3	5	4	4	3	Occasionally	Regularly	1	20-29	Male	South Asia	Bachelor's degree
3	3	3	3	4	4	3	3	3	3	3	3	Occasionally	Regularly	16+	50-59	Male	North America	Doctorate
4	4	4	4	5	4	4	5	4	5	4	4	Never	Occasionally	1	20-29	Female	North America	Bachelor's degree
5	5	5	5	5	5	5	4	5	5	5	5	Regularly	I try to watch every game	0	30-39	Male	North America	High school
3	2	4	3	2	3	3	3	3	4	2	3	Regularly	I try to watch every game	16+	20-29		North America	Some college
4	3	4	5	5	5	4	3	5	5	5	4	Regularly	I try to watch every game	16+	30-39	Female	North America	Bachelor's degree
3	4	3	4	4	4	3	2	3	4	3	3	Occasionally	Occasionally	2-5	20-29	Female	North America	Bachelor's degree
3	3	4	3	2	3	2	3	2	3	3	2	Regularly	I try to watch every game	2-5	20-29	Female	North America	Bachelor's degree
4	4	4	4	4	4	4	4	4	4	4	4	Rarely	Regularly	1	Under 20	Female	North America	High school
5	4	5	5	5	5	4	5	4	5	4	4	Regularly	Regularly	2-5	30-39	Male	North America	Some college
4	4	4	4	4	4	4	4	4	4	4	4	Regularly	I try to watch every game	2-5	30-39	Female	North America	Some college
4	4	4	4	4	4		4	4	4	4	4	Regularly	Regularly	2-5	30-39	Female	North America	Some college
4	4	5	4	3	4	3	3	4	5	3	3	Occasionally	I try to watch every game	16+	20-29	Female	North America	Some college
5	4	4	4	4	5	4	4	5	5	5	4	Rarely	I try to watch every game	1	30-39	Male	North America	Master's degree

(Continued from prior page)

Schedule Results League (Phase 1, part 2 of 2)

AssessmentID	Navigation and Findability	Structure	Identity	Substantiality	Completeness	Trustworthiness	Interaction	Customization	Savings	Confidence	Attractiveness	Enjoyment	How often have you used this app?	How often have you watched soccer games?	How many soccer game tickets have you bought this year?	Age	Gender	Geographic Region	Education
531	5	5	5	5	5	5	5	5	5	5	5	5	Regularly	I try to watch every game	2-5	50-59	Female	North America	High school
538	4	4	4	4	3	5	4	4	4	4	4	4	Occasionally	Regularly	6-9	30-39	Female	North America	Bachelor's degree
549	4	3	3	3	3	3	2	3	1	2	4	4	Regularly	I try to watch every game	16+	50-59		North America	Bachelor's degree
552	4	4	4	3	4	5	3	5	5	5	4	4	Almost every day	I try to watch every game	10-15	40-49	Female	North America	Bachelor's degree
579	4	3	3	4	4	4	3	2	3	4	2	2	Regularly	I try to watch every game	0	20-29	Male	North America	Bachelor's degree
585	2	2	3	3	4	4	1	3	1	2	1	2	Rarely	I try to watch every game	16+	30-39	Female	North America	Bachelor's degree
592	3	2	2	3	3	2	2	2	1	1	2	2	Rarely	Regularly	16+	50-59	Female	North America	Bachelor's degree
600	3	4	4	3	4	4	3	3	4	4	4	4	Rarely	Regularly	2-5	20-29	Female	North America	Some college
608	3	3	3	3	2	3	2	2	3	3	3	3	Occasionally	Occasionally	0	40-49	Female	North America	Master's degree
614	1	1	1	1	1	1	1	1	1	1	1	1	Occasionally	I try to watch every game	16+	30-39	Female	North America	Master's degree
644	3	3	3	3	3	3	1	3	3	4	2	2	Regularly	I try to watch every game	16+	30-39	Male	North America	Bachelor's degree
638	4	3	3	3	3	3	4	4	4	4	3	3	Occasionally	Occasionally	1	20-29	Male	North America	High school

Appendix L: Phase 2 Expert Ratings of Sounders FC iOS Application for Player Information and for Schedule, Results, and League

Player Info		Phase 4						
Criteria	**Values**	***Rater 1***	***Rater 2***	***Rater 3***	***Rater 4***	***Rater 5***	***Std Dev***	***Average***
Ease of Use	**Navigation and Findability** Browsing/searchability, Mediation, Orientation, Simplicity	4.00	4.00	**4.25**	4.00	**4.25**	0.14	4.10
Ease of Use	**Structure** Formatting/Presentation, Order, Accessibility	4.00	4.00	**4.50**	4.00	4.25	0.22	4.15
Noise Reduction	**Identity** Item Identification, Subject description, Summary, Selectivity, Precision	4.00	4.00	**3.50**	4.00	4.00	0.22	3.90
Noise Reduction	**Parsimony** Linkage/Referral, Order, Novelty	3.50	3.50	**3.00**	**4.50**	4.00	0.57	3.70
Quality	**Completeness** Accuracy, Comprehensiveness, Currency	**3.00**	3.50	3.50	**4.00**	**4.00**	0.42	3.60
Quality	**Trustworthiness** Reliability, Validity, Authority	**3.50**	**3.50**	4.00	4.00	4.00	0.27	3.80
Adaptability	**Interaction** Contextuality, Transaction, Feedback, Community/Social networking	**3.00**	3.50	3.50	**4.00**	3.50	0.35	3.50
Adaptability	**Customization** Flexibility, Simplicity, Trust, Individualization, Localization, Privacy	3.00	2.50	2.50	**2.00**	**3.50**	0.57	2.70
Performance	**Savings** Cost savings, Time savings	**3.50**	3.00	3.00	3.00	3.00	0.22	3.10
Performance	**Confidence** Security, Safety	3.50	**3.00**	**3.00**	**4.00**	3.50	0.42	3.40
Affection	**Attractiveness** Aesthetics, Satisfaction/rewarding/ incenting	4.00	**3.00**	3.50	**3.00**	**4.25**	0.57	3.55
Affection	**Enjoyment** Entertainment, Engagement, Stimulation	**4.00**	**3.00**	3.50	3.50	**4.00**	0.42	3.60
	Total averages	3.58	**3.38**	3.48	3.67	**3.85**	0.18	3.59

Schedule, Results and League		Phase 4						
Criteria	**Values**	***Rater 1***	***Rater 2***	***Rater 3***	***Rater 4***	***Rater 5***	***Std Dev***	***Average***
Ease of Use	**Navigation and Findability** Browsing/searchability, Mediation, Orientation, Simplicity	**4.00**	**4.00**	**4.00**	**4.00**	**4.00**	0.00	4.00
Ease of Use	**Structure** Formatting/Presentation, Order, Accessibility	4.00	4.00	4.00	**3.00**	4.00	0.45	3.80
Noise Reduction	**Identity** Item Identification, Subject description, Summary, Selectivity, Precision	**4.00**	**4.00**	**4.00**	**4.00**	**4.00**	0.00	4.00
Noise Reduction	**Parsimony** Linkage/Referral, Order, Novelty	3.50	**3.00**	**4.00**	**3.00**	**4.00**	0.50	3.50
Quality	**Completeness** Accuracy, Comprehensiveness, Currency	**4.50**	4.00	4.00	4.00	4.00	0.22	4.10
Quality	**Trustworthiness** Reliability, Validity, Authority	**4.00**	**4.00**	**4.00**	**4.00**	**4.00**	0.00	4.00
Adaptability	**Interaction** Contextuality, Transaction, Feedback, Community/Social networking	**2.50**	**4.00**	3.00	3.00	3.00	0.55	3.10
Adaptability	**Customization** Flexibility, Simplicity, Trust, Individualization, Localization, Privacy	**3.00**	**2.00**	2.50	**2.00**	**3.00**	0.50	2.50
Performance	**Savings** Cost savings, Time savings	**4.00**	**3.00**	**4.00**	**3.00**	3.50	0.50	3.50
Performance	**Confidence** Security, Safety	3.00	3.00	3.00	3.00	**3.50**	0.22	3.10
Affection	**Attractiveness** Aesthetics, Satisfaction/rewarding/ incenting	4.00	**3.00**	4.00	3.50	**4.25**	0.50	3.75
Affection	**Enjoyment** Entertainment, Engagement, Stimulation	4.00	**3.00**	4.00	4.00	4.25	0.49	3.85
	Total averages	3.71	**3.42**	3.71	**3.38**	**3.79**	0.19	3.60

Appendix M: Phase 4 User Ratings of Sounders FC iOS Application

Player Information (Phase 4, part 1 of 4)

AssessmentID	Navigation and Findability	Structure	Identity	Substantiality	Completeness	Trustworthiness	Interaction	Customization	Savings	Confidence	Attractiveness	Enjoyment	How often have you used this app?	How often have you watched soccer games?	How many soccer game tickets have you bought this year?	Age	Gender	Geographic Region	Education
700	3	3	4	3	3	4	3	2	3	5	4	3	Almost every day	I try to watch every game	16+	50-59	Female	North America	Some college
697	3	2	3	3	3	4	2	1	3	3	3	3	Regularly	I try to watch every game	16+	40-49	Male	North America	Bachelor's degree
684	4	4	4	5	4	4	5	5	3	4	4	4	Rarely	Regularly	2-5	30-39	Female	North America	Bachelor's degree
681	3	4	4	3	4	3	3	4	3	4	4	3	Regularly	Regularly	10-15	40-49	Female	North America	Bachelor's degree
717	4	5	5	4	5	4	4	5	4	4	5	5	Occasionally	Regularly	2-5	Under 20	Male	Europe (except R	High school
674	4	4	2	4	4	3	2	2	2	3	4	3	Regularly	I try to watch every game	1	40-49	Male	North America	Master's degree
685	3	3	3	3	4	3	3	3	2	4	3	2	Regularly	I try to watch every game	16+	40-49	Female	North America	Master's degree
687	3	3	5	4	4	5	4	2	3	3	2	2	Never	I try to watch every game	16+	30-39	Male	North America	Master's degree
688	4	5	5	4	3	5	2	1	5	5	5	5	Almost every day	I try to watch every game	16+	20-29	Male	North America	Master's degree
695	2	3	2	4	4	4	2	3	3	4	3	2	Almost every day	I try to watch every game	16+	20-29	Female	North America	Master's degree
704	3	3	4	2	4	3	3	2	2	3	3	3	Regularly	I try to watch every game	16+	30-39	Male	North America	Bachelor's degree
706	3	2	2	2	3	3	3	3	2	2	1	2	Occasionally	Regularly	16+	30-39	Male	North America	Master's degree
744	3	3	2	3	4	3	3	4	3	4	2	1	Regularly	I try to watch every game	16+	20-29	Male	North America	Some college
729	3	3	3	2	3	4	2	2	2	3	3	2	Regularly	I try to watch every game	16+	40-49	Male	North America	Bachelor's degree
719	4	4	4	5	4	5	4	5	5	5	5	4	Occasionally	Regularly	0	Under 20	Female	North America	Some college
716	3	4	3	3	2	4	2	3	3	3	2	4	Regularly	I try to watch every game	16+	40-49	Male	North America	Bachelor's degree
713	4	4	4	3	3	4	2	4	4	5	5	4	Occasionally	I try to watch every game	6-9	30-39	Male	North America	Some college
733	5	4	4	4	5		4	4	4	5	5	4	Occasionally	I try to watch every game	6-9	20-29	Female	North America	Bachelor's degree
756	3	4	4	3	4								Almost every day	I try to watch every game	10-15	30-39	Male	North America	Master's degree
737	4	5	5	5	5	4	4	5	5	5	4	5	Regularly	Regularly	16+	30-39	Male	North America	Bachelor's degree
739	2	2	3	4	3	5	2	2	2	4	4	3	Regularly	Regularly	16+	50-59	Male	North America	Bachelor's degree

(Continued on next page)

(Continued from prior page)

Player Information (Phase 4, part 2 of 4)

AssessmentID	Navigation and Findability	Structure	Identity	Substantiality	Completeness	Trustworthiness	Interaction	Customization	Savings	Confidence	Attractiveness	Enjoyment	How often have you used this app?	How often have you watched soccer games?	How many soccer game tickets have you bought this year?	Age	Gender	Geographic Region	Education
741	5	5	4	3	5	4	5	4	3	4	5	5	Regularly	Occasionally	2-5	20-29	Male	North America	Some college
765	3	3	3	2	3	4	3	3	4	4	4	4	Occasionally	I try to watch every game	6-9	30-39	Male	North America	Some college
747	4	4	4	4	4	4	4	4	4	4	4	4	Regularly	Regularly	0	40-49	Male	North America	Some college
753	4	3	4	3	3	3	3	4	3	4	4	3	Regularly	I try to watch every game	16+	40-49	Male	North America	Bachelor's degree
767	4	4	4	4	4	4	4	4	4	4	4	4	Regularly	I try to watch every game	6-9	20-29	Female	North America	Bachelor's degree
771	5	5	5	5	5	5	5	5	5	5	5	5	Regularly	I try to watch every game	10-15	20-29	Female	North America	Some college
786	3	4	3	2	2	2	1	1	2	4	5	2	Regularly	I try to watch every game	16+	20-29	Male	North America	Some college
856	3	3	3	1	2	2	2	3					Regularly	Regularly	6-9	30-39	Female	North America	Master's degree
848	4	4	4	4	4	4	4	4	4	4	4	4	Occasionally	Regularly	2-5	30-39	Female	North America	Some college
772	4	4	4	4	4	4	4	4	4	4	4	4	Rarely	I try to watch every game	0	40-49	Female	North America	Doctorate
775	2	3	3	1	2	2	2	3	2	3	4	2	Regularly	Regularly	16+	30-39	Male	North America	Bachelor's degree
820	5	5	5	4	5	5	3	4	3	4	5	4	Rarely	I try to watch every game	0	30-39	Male	North America	High school
805	3	3	4	3	4	4	4		3			4	Occasionally	I try to watch every game	16+	20-29	Female	North America	Bachelor's degree
807	4	4	4	3	3	4	3	3	3	3	3	3	Regularly	I try to watch every game	16+	50-59	Female	North America	Master's degree
798	5	4	5	4	3	4	4	3	4	3	3	3	Regularly	I try to watch every game	16+	30-39	Male	North America	Bachelor's degree
813	4	4	4	4	5	4	3	4	3	4	4	4	Occasionally	Regularly	10-15	20-29	Female	North America	Bachelor's degree
815	5	5	5	5	5	5							Occasionally	I try to watch every game	2-5	20-29	Male	North America	Bachelor's degree
824	4	4	3	4	4	4	3	3	4	4	4	4	Regularly	I try to watch every game	10-15	30-39	Female	North America	Some college
833	4	4	5	5	4	5	2	2	3	4	4	4	Never	Regularly	0	30-39	Male	east Asia (except I	Bachelor's degree

(Continued on next page)

(Continued from prior page)

Player Information (Phase 4, part 3 of 4)

AssessmentID	Navigation and Findability	Structure	Identity	Substantiality	Completeness	Trustworthiness	Interaction	Customization	Savings	Confidence	Attractiveness	Enjoyment	How often have you used this app?	How often have you watched soccer games?	How many soccer game tickets have you bought this year?	Age	Gender	Geographic Region	Education
834	3	4	5	4	5	4	4	5	5	4	5	3	Never	Regularly	1	50-59	Male	North America	Bachelor's degree
846	5	5	5	5	5	5	5	5	5	5	5	5	Almost every day	Regularly	2-5	30-39	Male	North America	High school
842	3	4	4	2	1	3	2	3	2	3	1	1	Regularly	I try to watch every game	16+	30-39	Male	North America	Some college
838	4	4	4	4	4	5	5	3	4	5	5	5	Never	Occasionally	0	40-49	Female	North America	Bachelor's degree
854	3	3	4	4	3	4	4	3	3	2	3	3	Regularly	I try to watch every game	16+	30-39	Male	North America	Some college
851	3	4	4	3	4	5	4	4	4	5	5	4	Regularly	Regularly	16+	30-39	Male	North America	Bachelor's degree
870	4	4	5	4	5	5	4	4	5	4	5	4	Regularly	I try to watch every game	6-9	40-49	Female	North America	Master's degree
858	3	3	4	2	3	4	3	4	4	3	4	3	Rarely	I try to watch every game	6-9	30-39	Female	North America	Bachelor's degree
877	5	5	4	4	5	4	4	3	5	5	5	4	Never	I try to watch every game	0	50-59	Male	North America	Bachelor's degree
868	4	4	3	5	4	4	5	3	4	5	4	4	Regularly	I try to watch every game	16+	50-59	Female	North America	Master's degree
880	4	4	4	4	4	4		4	4	4	4	4	Occasionally	Occasionally	6-9	40-49	Male	North America	Master's degree
873	4	4	5	4	5	5	3	4	5	5	4	4	Occasionally	I try to watch every game	2-5	20-29	Female	North America	Bachelor's degree
881	3	4	3	3	4	5	4	1	3	4	5	3	Regularly	I try to watch every game	16+	30-39	Male	North America	Bachelor's degree
930	3	2	3	4	4	3	2	3	2	3	4	2	Regularly	I try to watch every game	16+	30-39	Female	North America	Bachelor's degree
908	4	3	4	3	5	5	3	4	4	5	4	4	Occasionally	I try to watch every game	6-9	30-39	Male	North America	Bachelor's degree
892	4	4	4	4	4	4	4	4	4	4	5	5	Rarely	Regularly	2-5	30-39	Female	North America	Some college
887	3	3	3	4	3	4	1	2	3	4	3	4	Occasionally	Regularly	16+	40-49	Female	North America	Master's degree
889	4	4	4	5	5	5	3	4	5	5	5	4	Occasionally	I try to watch every game	16+	40-49	Female	North America	Bachelor's degree
922	4		5	3	5	5	3	1	5	4	4		Occasionally	I try to watch every game	16+	40-49	Male	North America	Master's degree

(Continued on next page)

(Continued from prior page)

Player Information (Phase 4, part 4 of 4)

AssessmentID	Navigation and Findability	Structure	Identity	Substantiality	Completeness	Trustworthiness	Interaction	Customization	Savings	Confidence	Attractiveness	Enjoyment	How often have you used this app?	How often have you watched soccer games?	How many soccer game tickets have you bought this year?	Age	Gender	Geographic Region	Education
897	2	4	3	4	4	4	2	4	2	4	5	5	Regularly	I try to watch every game	10-15	20-29	Male	North America	High school
899	4	4	4	4	3	5	4	4	4	5	4	4	Occasionally	Regularly	2-5	30-39	Male	North America	Some college
901	4	4	4	3	3	4	2	3	4	4	5	4	Regularly	Regularly	0	40-49	Female	North America	Bachelor's degree
912	5	4	4	5	4	5	4	5	4	5	5	5	Occasionally	I try to watch every game	1	Under 20	Male	North America	Some college
928	3	4	4	5	4	4	3	3	5	5	5	5	Never	Regularly	6-9	40-49	Female	North America	Master's degree
927	5	5	5	5	5	5	5	5	5	5	5	5	Regularly	I try to watch every game	1	40-49	Male	North America	Bachelor's degree
943	1	2	5	2	2	5	1	3	3	3	1	1	Occasionally	I try to watch every game	6-9	30-39	Male	North America	Master's degree
937	4	4	4	4	4	4	3	4	4	5	5	5	Occasionally	Regularly	0	50-59	Female	North America	Some college
939	4	4	4	3	4	4	3	4	4	4	4	4	Regularly	Regularly	16+	30-39	Male	North America	Some college
941	4	4	4	5	4	5	3	4	5	4	5	4	Never	Regularly	16+	50-59	Male	North America	Some college
946	5	4	5	5	4	4	5	5	5	5	5	5	Regularly	I try to watch every game	16+	30-39	Female	North America	Some college
948	4	4	5	3	4	4	3	4	4	4	4	4	Occasionally	I try to watch every game	10-15	20-29	Female	North America	High school
959	5	4	4	4	4	4	4	5	5	5	5	4	Never	I try to watch every game	16+	30-39	Female	North America	Master's degree
957	4	5	4	4	4	5	3	4	5	4	5	4	Occasionally	Occasionally	0	20-29	Female	North America	Bachelor's degree
967		5	5	5	4	5	3	3	4	5	5	5	Regularly	I try to watch every game	10-15	20-29	Male	North America	Bachelor's degree
963	5	5	5	5	5	5	5	5	5	5	5	5	Almost every day	Regularly	0	50-59	Male	North America	Bachelor's degree
970	4	5	5	4	5	5	4	5	4	5	5	5	Occasionally	Regularly	2-5	40-49	Female	North America	Bachelor's degree

Schedule Results League (Phase 4, part 1 of 2)

AssessmentID	Navigation and Findability	Structure	Identity	Substantiality	Completeness	Trustworthiness	Interaction	Customization	Savings	Confidence	Attractiveness	Enjoyment	How often have you used this app?	How often have you watched soccer games?	How many soccer game tickets have you bought this year?	Age	Gender	Geographic Region	Education
682	3	3	4	3	3	4	3	4	3	4	4	4	Regularly	Regularly	10-15	40-49	Female	North America	Bachelor's degree
701	3	3	3	3	3	3	2	2	3	4	3	3	Almost every day	I try to watch every game	16+	50-59	Female	North America	Some college
689	5	4	3	3	4	5	4	4	3	2	1	1	Almost every day	I try to watch every game	16+	20-29	Male	North America	Master's degree
730	3	3	3	2	3	3	2	2	2	3	3	3	Regularly	I try to watch every game	16+	40-49	Male	North America	Bachelor's degree
707	3	4	3	3	3	4	2	2	2	2	1	1	Occasionally	Regularly	16+	30-39	Male	North America	Master's degree
718	4	4	4	5	5	4	4	5	4	5	5	5	Occasionally	Regularly	2-5	Under 20	Male	rope (except Rus	High school
714	4	4	4	4	4	4	3	4	4	5	5	5	Occasionally	I try to watch every game	6-9	30-39	Male	North America	Some college
720	5	5	4	5	5	5	4	4	5	5	5	4	Occasionally	Regularly	0	Under 20	Female	North America	Some college
745	2	3	3	3	3	4	3	4	2	3	2	1	Regularly	I try to watch every game	16+	20-29	Male	North America	Some college
736	2	4	3	2	4	4	2	3	4	5	4	4	Almost every day	I try to watch every game	16+	30-39	Male	North America	Bachelor's degree
738	5	5	5	5	5	5	5	5	5	5	5	5	Regularly	Regularly	16+	30-39	Male	North America	Bachelor's degree
754	5	5	5	5	5	5	5	5	5	5	5	4	Regularly	I try to watch every game	16+	40-49	Male	North America	Bachelor's degree
768	4	4	4	4	4	4	4	4	4	4	4	4	Regularly	I try to watch every game	6-9	20-29	Female	North America	Bachelor's degree
773	4	4	4	4	4	4	3	4	4	4	4	4	Rarely	I try to watch every game	0	40-49	Female	North America	Doctorate
804	2	2	3	3	3								Occasionally	I try to watch every game	16+	30-39	Female	North America	Bachelor's degree
787	2	1	1	4	1	1	1	1	4	4	5	2	Regularly	I try to watch every game	16+	20-29	Male	North America	Some college
821	4	4	4	4	4	4	3	4	4	5	5	4	Rarely	I try to watch every game	0	30-39	Male	North America	High school
847	5	5	5	5	5	5	5	5	5	5	5	5	Almost every day	Regularly	2-5	30-39	Male	North America	High school
825	4	4	4	4	4	4	3	3	4	4	4	4	Regularly	I try to watch every game	10-15	30-39	Female	North America	Some college
827	3	3	3	1	1	2	2	3	3	3	2	2	Regularly	I try to watch every game	16+	40-49	Male	North America	Bachelor's degree
843	2	2	3	2	2	3	2	3	1	2	2	1	Regularly	I try to watch every game	16+	30-39	Male	North America	Some college

(Continued on next page)

(Continued from prior page)

Schedule Results League (Phase 4, part 2 of 2)

AssessmentID	Navigation and Findability	Structure	Identity	Substantiality	Completeness	Trustworthiness	Interaction	Customization	Savings	Confidence	Attractiveness	Enjoyment	How often have you used this app?	How often have you watched soccer games?	How many soccer game tickets have you bought this year?	Age	Gender	Geographic Region	Education
836	5	5	5	5	5	5	3	5	5	5	5	5	Never	Regularly	0	30-39	Male	ast Asia (except	Bachelor's degree
852	4	4	5	4	5	5	4	4	5	5	5	4	Regularly	Regularly	16+	30-39	Male	North America	Bachelor's degree
859	4	2	3	4	2	3	4	4	4	4	4	3	Rarely	I try to watch every game	6-9	30-39	Female	North America	Bachelor's degree
869	5	4	5	5	5	5	5	4	5	5	5	4	Regularly	I try to watch every game	16+	50-59	Female	North America	Master's degree
878	5	5	5	4	5	5	5	5	5	5	5	5	Never	I try to watch every game	0	50-59	Male	North America	Bachelor's degree
874	5	4	5	5	4	5	3	2	5	4	5	4	Occasionally	Occasionally	1	20-29	Male	North America	Some college
879	4	4	4	4	4	4	0	2	4	5	4	4	Occasionally	Occasionally	6-9	40-49	Male	North America	Master's degree
882	4	3	4	4	5	4	3	3	4	5	4	4	Regularly	I try to watch every game	16+	30-39	Male	North America	Bachelor's degree
913	5	4	4	5	4	4	4	5	5	5	5	4	Occasionally	I try to watch every game	1	Under 20	Male	North America	Some college
888	4	4	5	5	5	5	1	1	4	5	4	5	Occasionally	Regularly	16+	40-49	Female	North America	Master's degree
898	3	4	4	3	4	4	2	3	4	4	2	5	Regularly	I try to watch every game	10-15	20-29	Male	North America	High school
900	4	3	3	4	5	5	4	4	4	4	4	4	Occasionally	Regularly	2-5	30-39	Male	North America	Some college
902	3	3	3	3	4	4	2	3	4	4	4	4	Regularly	Regularly	0	40-49	Female	North America	Bachelor's degree
931	3	4	4	2	3	4	2	3	3	3	5	3	Regularly	I try to watch every game	16+	30-39	Female	North America	Bachelor's degree
938	5	4	4	4	4	4	4	4	4	5	5	5	Occasionally	Regularly	0	50-59	Female	North America	Some college
947	5	5	5	5	5	5	5	5	5	5	5	5	Regularly	I try to watch every game	16+	30-39	Female	North America	Some college
949	4	4	4	4	4	5	4	3	3	4	4	0	Occasionally	I try to watch every game	10-15	20-29	Female	North America	High school
958	4	4	5	4	4	5	3	4	4	4	5	5	Occasionally	Occasionally	0	20-29	Female	North America	Bachelor's degree
960	5	5	5	5	4	4	4	5	5	5	5	5	Never	I try to watch every game	16+	30-39	Female	North America	Master's degree
968	5	4	4	4	5	4	5	4		5		5	Regularly	I try to watch every game	10-15	20-29	Male	North America	Bachelor's degree
971	5	5	4	5	5	5	4	4	5	5	5	5	Occasionally	Regularly	2-5	40-49	Female	North America	Bachelor's degree

Appendix N: Phase 1 and Phase 2 Expert Rating Comparisons for Remaining Six Scenarios

Appendix N: *Team News - Part 1 of 6*

Phase 1 and 2 Ratings for Team News			
Values	***Phase 1 Expert***	***Phase 2 Expert***	***Difference***
Navigation and Findability	3.30	3.60	0.30
Structure	3.80	3.80	0.00
Identity	3.70	3.80	0.10
Parsimony	3.40	3.50	0.10
Completeness	3.80	4.00	0.20
Trustworthiness	3.80	3.90	0.10
Interaction	2.70	3.10	0.40
Customization	2.30	2.30	0.00
Savings	3.10	3.30	0.20
Confidence	3.20	3.20	0.00
Attractiveness	3.80	4.00	0.20
Enjoyment	3.40	3.60	0.20
Total averages	3.36	3.51	0.15

Appendix N: Merchandise and Store - Part 2 of 6

Phase 1 and 2 Ratings for Merchandise and Store

Values	Phase 1 Expert	Phase 2 Expert	Difference
Navigation and Findability	3.30	3.30	0.00
Structure	3.10	3.40	0.30
Identity	3.50	3.60	0.10
Parsimony	3.00	3.10	0.10
Completeness	3.50	3.60	0.10
Trustworthiness	3.50	3.60	0.10
Interaction	3.30	3.40	0.10
Customization	3.00	3.30	0.30
Savings	3.20	3.30	0.10
Confidence	3.10	3.10	0.00
Attractiveness	3.00	3.20	0.20
Enjoyment	2.90	3.10	0.20
Total averages	3.20	3.33	0.13

Appendix N: Ticketing - Part 3 of 6

Phase 1 and 2 Ratings for Ticketing			
Values	***Phase 1 Expert***	***Phase 2 Expert***	***Difference***
Navigation and Findability	1.20	3.30	2.10
Structure	1.20	3.40	2.20
Identity	1.20	3.40	2.20
Parsimony	1.20	3.10	1.90
Completeness	1.20	3.50	2.30
Trustworthiness	1.20	3.60	2.40
Interaction	1.20	3.10	1.90
Customization	1.20	2.90	1.70
Savings	1.20	3.20	2.00
Confidence	1.20	2.80	1.60
Attractiveness	1.20	3.10	1.90
Enjoyment	1.20	2.80	1.60
Total averages	1.20	3.18	1.98

Appendix N: Social Media Integration - Part 4 of 6

Phase 1 and 2 Ratings for Social Media Integration

Values	Phase 1 Expert	Phase 2 Expert	Difference
Navigation and Findability	2.70	3.00	0.30
Structure	2.60	3.10	0.50
Identity	2.80	3.10	0.30
Parsimony	2.60	2.90	0.30
Completeness	2.70	2.90	0.20
Trustworthiness	2.90	3.10	0.20
Interaction	2.70	2.90	0.20
Customization	2.60	2.60	0.00
Savings	2.80	3.10	0.30
Confidence	2.90	2.90	0.00
Attractiveness	2.80	3.00	0.20
Enjoyment	2.80	3.10	0.30
Total averages	2.74	2.98	0.23

Appendix N: Game Day and Stadium Guide - Part 5 of 6

Phase 1 and 2 Ratings Game Day and Stadium Guide

Values	*Phase 1 Expert*	*Phase 2 Expert*	*Difference*
Navigation and Findability	1.00	2.80	1.80
Structure	1.00	2.60	1.60
Identity	1.00	2.50	1.50
Parsimony	1.00	2.50	1.50
Completeness	1.00	2.40	1.40
Trustworthiness	1.00	2.90	1.90
Interaction	1.00	2.20	1.20
Customization	1.00	2.30	1.30
Savings	1.00	2.50	1.50
Confidence	1.00	2.90	1.90
Attractiveness	1.00	2.70	1.70
Enjoyment	1.00	2.60	1.60
Total averages	1.00	2.58	1.58

Appendix N: In-Seat Concessions - Part 6 of 6

Phase 1 and 2 Ratings for In-Seat Concessions			
Values	***Phase 1 Expert***	***Phase 2 Expert***	***Difference***
Navigation and Findability	1.00	1.00	0.00
Structure	1.00	1.00	0.00
Identity	1.00	1.00	0.00
Parsimony	1.00	1.00	0.00
Completeness	1.00	1.00	0.00
Trustworthiness	1.00	1.00	0.00
Interaction	1.00	1.00	0.00
Customization	1.00	1.00	0.00
Savings	1.00	1.00	0.00
Confidence	1.00	1.00	0.00
Attractiveness	1.00	1.00	0.00
Enjoyment	1.00	1.00	0.00
Total averages	1.00	1.00	0.00

Appendix O: Development Items in the Sounders FC iOS Application Backlog

Item	Version	Work Item	Scenario	Cluster
1	3.5.5	Updated Match Details screen	Schedule, Results and League	Structure, Completeness, Attractiveness, Enjoyment
2	3.5.5	Updated MatchDay Guide screen	Schedule, Results and League	Structure, Attractiveness
3	3.5.5	Various UI adjustments and bug fixes	TBD	TBD
4	3.5.3 (1809)	Standings layout updated to include the Supporters' Shield	Schedule, Results and League	Navigation and Findability, Structure, Completeness, Savings, Enjoyment
5	3.5.3 (1809)	Fixed bug related to news push notifications	TeamNews	Completeness, Savings
6	3.5.3 (1809)	Changed layout of the news item cell.	TeamNews	Navigation and Findability, Structure, Attractiveness
7	3.5.3 (1809)	More high resolution images.	TBD	Identity, Completeness, Attractiveness, Enjoyment
8	3.5.3 (1809)	Scroll to top on status bar touch!	Schedule, Results and League, Player Info, Team News	Navigation and Findability, Structure, Parsimony, Savings
9	3.5.3 (1809)	Various UI adjustments and bug fixes.	TBD	TBD
10	3.5.2 (1808)	Support for roster and news item push notifications.	TBD	TBD
11	3.5.2 (1808)	News article sharing.	TeamNews	Structure, Interaction, Savings

12	3.5.2 (1808)	Fixed incorrect score issue on Live Scores reported by a Twitter user. (Thx!)	Schedule, Results and League	Completeness, Trustworthiness
13	3.5.2 (1808)	Fixed past match scores to include shootouts.	Schedule, Results and League	Completeness, Trustworthiness
14	3.5.2 (1808)	Fixed bug where all past scores appear as draws in certain circumstances.	Schedule, Results and League	Completeness, Trustworthiness
15	3.5.2 (1808)	Various UI adjustments and bug fixes.	TBD	TBD
16	3.5.1 (1807)	Redesigned Vote for Man of the Match screen.	Schedule, Results and League	Structure, Interaction, Attractiveness, Enjoyment
17	3.5.1 (1807)	Redesigned Standings screen.	Schedule, Results and League	Navigation and Findability, Structure, Identity, Parsimony, Savings,
18	3.5.1 (1807)	Fixed jumpiness in News, Video, Twitter pull to refresh.	TeamNews	Structure, Attractiveness, Enjoyment
19	3.5.1 (1807)	Fixed multiple calls to table view reload on News & Video screen.	TeamNews	Trustworthiness, Enjoyment
20	3.5.1 (1807)	Images downloaded from internet use disk cache.	TBD	Completeness, Enjoyment
21	3.5.1 (1807)	Fade in animations for images on Roster, Schedule and News.	Schedule, Results and League	Structure, Attractiveness, Enjoyment
22	3.5.1 (1807)	Higher resolution Sounders logos for @3x screens.	TBD	Structure, Attractiveness
23	3.5	Support for season ticket renewals	Ticketing	Navigation and Findability, Identity, Completeness

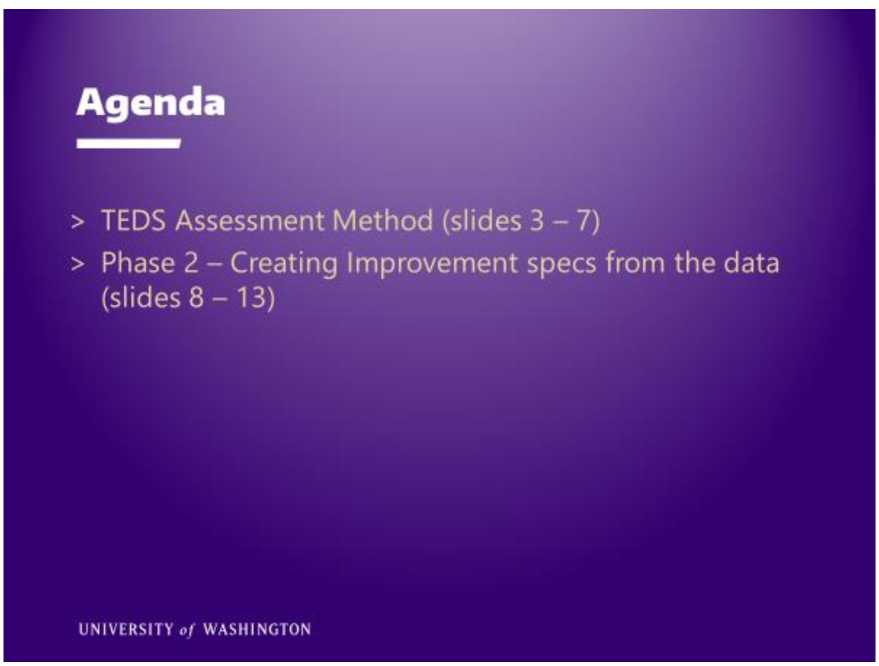
24	3.5.1 (1807)	Various UI adjustments and bug fixes.	TBD	Attractiveness
25	3.5 (1806)	Please test everything. There have been many UI adjustments	TBD	Attractiveness
26	3.4 (1805)	Support for native resolutions on iPhone 6 and iPhone 6 Plus (and up).	TBD	Trustworthiness
27	3.3 (Jun 20, 2016)	Starting XI now available in the app sooner	Game Day and Stadium Guide PlayerInfo	Navigation and Findability, Identity, Completeness, Enjoyment
28	3.3 (Jun 20, 2016)	Starting XI now added to Match Ticker	Game Day and Stadium Guide	Navigation and Findability, Identity, Completeness, Enjoyment
29	3.3 (Jun 20, 2016)	Scores for other matches currently in progress now available - sponsored by T-	Schedule, Results and League	Completeness, Savings, Enjoyment
30	3.3 (Jun 20, 2016)	Certain Promotions are now redeemable using the app	Game Day and Stadium Guide	Savings
31	3.3 (Jun 20, 2016)	Fixed issues with display of header bar buttons/icons when using Ticketmaster	Ticketing	Structure, Attractiveness
32	3.3 (Jun 20, 2016)	Testing updates to Player Statistics views	PlayerInfo	Attractiveness, Enjoyment
33	3.3 (Jun 20, 2016)	Fixed scroll jumping when in Schedule, Roster, or News/Feed views	PlayerInfo Schedule, Results and League	Trustworthiness, Confidence, Attractiveness, Enjoyment
34	3.3 (Jun 20, 2016)	Fixed bugs and crashes reported by users	TBD	Trustworthiness, Confidence
35	3.2 (Apr 14, 2016)	Improvements for the Matchday Program	Game Day and Stadium Guide	Completeness, Savings, Enjoyment

36	3.2 (Apr 14, 2016)	Improvements for live audio stream - details about start time - allow to play past	TeamNews	Completeness, Confidence, Enjoyment
37	3.2 (Apr 14, 2016)	Add archived audio feeds from prior matches (accessible through the	TeamNews	Completeness, Savings, Enjoyment
38	3.2 (Apr 14, 2016)	Ability to handle MatchPass BuyNow rewards in app	Ticketing Game Day and Stadium Guide	Completeness, Savings
39	3.2 (Apr 14, 2016)	Ability to handle match sponsors	Game Day and Stadium Guide	Interaction, Attractiveness
40	3.2 (Apr 14, 2016)	Updated version of Ticketmaster SDK	Ticketing	Completeness
41	3.2 (Apr 14, 2016)	Bug fixes from user feedback	TBD	TBD
42	3.1 (Mar 16, 2016)	Corrected display of Events, Live Feed, and implemented Match Stats for past matches	TeamNews Game Day and Stadium Guide	Completeness, Trustworthiness, Savings, Confidence, Attractiveness,
43	3.1 (Mar 16, 2016)	Profile more clearly labeled MatchPass Profile	Ticketing	Identity, Customization
44	3.1 (Mar 16, 2016)	Added Next button to sign in/sign up to make email entry clearer	Ticketing	Identity, Customization
45	3.1 (Mar 16, 2016)	Icons removed from main menu per user feedback	TBD	Structure, Parsimony
46	3.1 (Mar 16, 2016)	Data refresh enabled when app returns to foreground after > 1 hour of inactivity	TeamNews	Completeness, Trustworthiness, Savings, Confidence, Enjoyment
47	3.1 (Mar 16, 2016)	MatchDay Program made available from past match information in Schedule	Schedule, Results and League Game Day and Stadium	Completeness, Trustworthiness

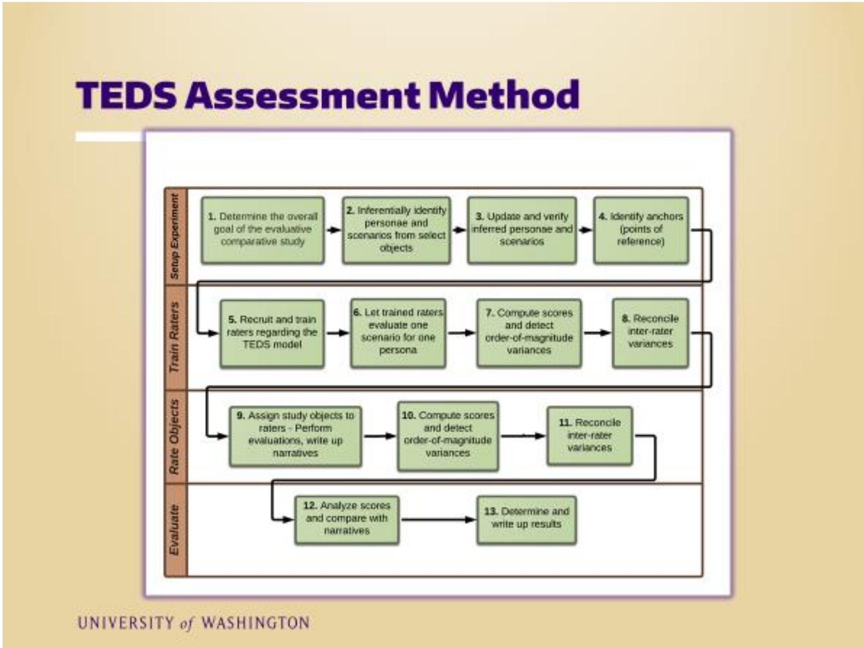
48	3.1 (Mar 16, 2016)	Adjusted some layout issues for smaller phones	TBD	Structure, Attractiveness
49	3.1 (Mar 16, 2016)	Adjusted layout of offer/promo images within roster and schedule	Schedule, Results and League	Structure, Attractiveness
50	3.1 (Mar 16, 2016)	Correctly placed "Designated" label above player headshot in details	PlayerInfo	Identity, Attractiveness
51	3.1 (Mar 16, 2016)	Showcase thumbnail image of MatchDay program on home page and on past-	Game Day and Stadium Guide	Navigation and Findability, Structure, Attractiveness
52	3.1 (Mar 16, 2016)	Fix live feed details on past-match	Game Day and Stadium Guide	Completeness, Trustworthiness, Savings, Enjoyment
53	3.1 (Mar 16, 2016)	Remove unused Social feature, replace with detailed match stats	Social Media Integration Schedule, Results and League	Structure, Parsimony, Completeness, Enjoyment
54	3.1 (Mar 16, 2016)	Fixed ticket links for away matches	Ticketing	Navigation and Findability, Trustworthiness, Savings
55	3.0 (Feb 16, 2016)	More content across MLS for the first team and opponents, schedule and	PlayerInfo Schedule, Results and League	Completeness, Savings, Enjoyment
56	3.0 (Feb 16, 2016)	Mobile ticketing and Ticketmaster integration.	Ticketing	Completeness, Savings
57	3.0 (Feb 16, 2016)	Access late-breaking offers and promotions.	Game Day and Stadium Guide	Completeness, Savings, Enjoyment

Appendix P: TEDS Delphi Participant Training PowerPoint Presentation

Pages 1 & 2



Pages 3 & 4

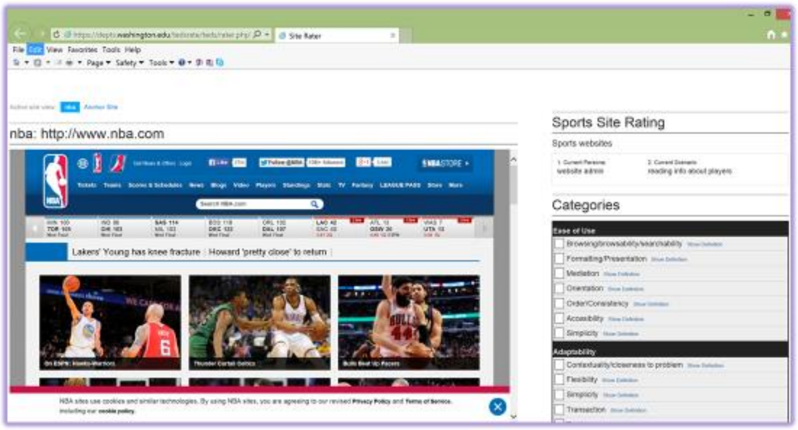


TEDS Assessment Method
Setup Experiment
1. Determine the overall goal of the evaluative comparative study
2. Inferentially identify personae and scenarios from select objects
3. Update and verify inferred personae and scenarios
4. Identify anchors (points of reference)
Train Raters
5. Recruit and train raters regarding the TEDS model
6. Let trained raters evaluate one scenario for one persona
7. Compute scores and detect order-of-magnitude variances
8. Reconcile inter-rater variances
Rate Objects
9. Assign study objects to raters - Perform evaluations, write up narratives
10. Compute scores and detect order-of-magnitude variances
11. Reconcile inter-rater variances
Evaluate
12. Analyze scores and compare with narratives
13. Determine and write up results
UNIVERSITY of WASHINGTON

Soccer Website Use Scenarios
• Player Information
• Team News
• Schedule, Results and League
• Merchandise and Store
• Ticketing
• Social Media Integration
• Game Day and Stadium Guide
• In-Seat Concessions
UNIVERSITY of WASHINGTON

Pages 5 & 6

TEDS Assessment Method Tool



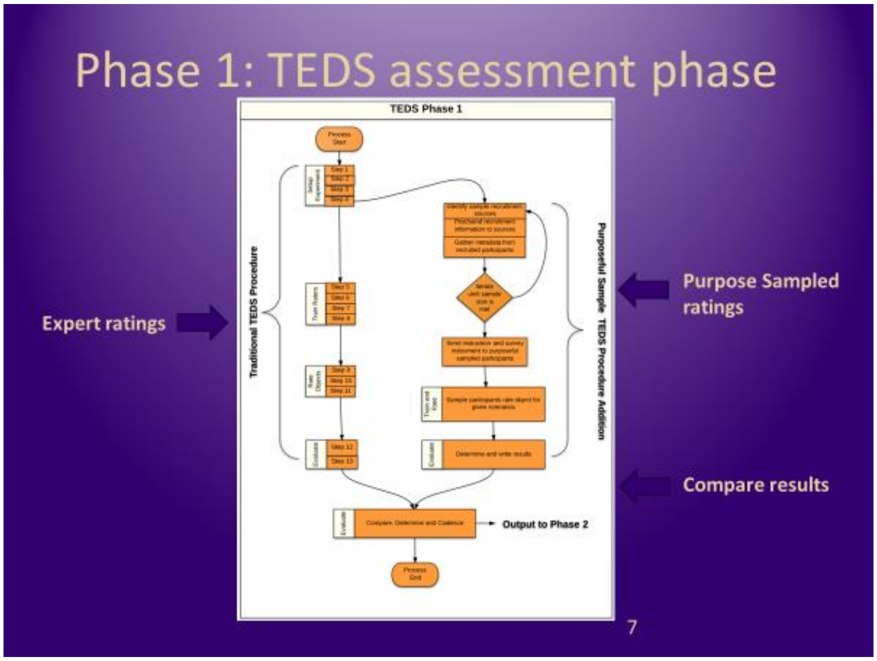
UNIVERSITY *of* WASHINGTON

Example Comparison – Player Information

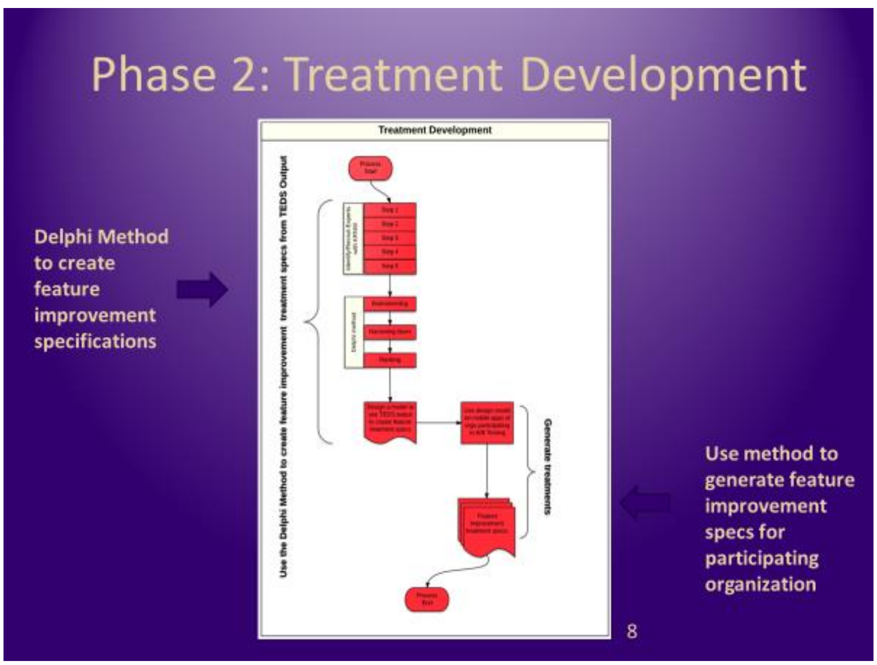
Values	…hawks	Barca	Sounders	Real Madrid	Chicago Fire	Bayern	OneFootball	ESPN FC	Brooklyn Nets	Arsenal FC	Sounders Beta	Average
Navigation and Findability Browsing/searchability, Mediation, Orientation, Simplicity	3.65	3.70	3.10	3.70	3.20	3.50	3.90	1.90	3.80	3.50	4.00	3.45
Structure Formatting/Presentation, Order, Accessibility	3.60	4.00	3.10	3.80	3.10	3.00	3.90	2.30	3.90	3.50	3.60	3.44
Identity Item Identification, Subject description, Summary, Selectivity, Precision	3.64	3.60	2.90	4.00	3.40	3.00	3.80	2.40	3.70	3.40	3.50	3.39
Parsimony Linkage/Referral, Order, Novelty	3.33	3.80	2.90	3.50	2.90	2.60	3.60	2.10	2.80	3.00	3.60	3.10
Completeness Accuracy, Comprehensiveness, Currency	3.87	3.40	1.70	4.00	2.90	3.40	3.90	2.40	3.68	3.40	3.20	3.26
Trustworthiness Reliability, Validity, Authority	4.67	4.20	2.80	3.90	3.40	3.70	3.80	2.30	3.60	3.70	3.70	3.62
Interaction Contextuality, Transaction, Feedback, Community/Social networking	2.70	3.70	1.80	3.80	1.90	2.90	3.60	1.80	2.50	2.70	3.30	2.79
Customization Flexibility, Simplicity, Trust, Individualization, Localization, Privacy	2.93	2.80	2.20	3.10	2.30	2.40	3.30	2.00	2.30	2.20	2.60	2.56
Savings Cost savings, Time savings	3.20	3.20	2.78	3.80	2.60	3.00	3.20	2.00	3.60	2.70	3.00	3.01
Confidence Security, Safety	3.30	3.30	2.90	3.40	2.60	3.40	3.40	2.20	3.30	3.20	3.30	3.12
Attractiveness Aesthetics, Satisfaction/rewarding/ incenting	3.30	3.80	2.60	4.40	2.60	2.80	3.70	2.40	3.70	3.30	3.20	3.25
Enjoyment Entertainment, Engagement, Stimulation	3.27	3.40	2.30	4.10	2.60	2.80	3.60	2.20	3.40	3.00	3.20	3.08
Average Likert	3.45	3.58	2.59	3.79	2.79	3.04	3.64	2.17	3.36	3.13	3.35	3.17

UNIVERSITY *of* WASHINGTON

Pages 7 & 8



Phase 1: TEDS assessment phase
TEDS Phase 1
Traditional TEDS Procedure
Purposeful Sample TEDS Procedure Addition
Output to Phase 2
Expert ratings
Purpose Sampled ratings
Compare results
7



Phase 2: Treatment Development
Treatment Development
Use the Delphi Method to create feature improvement treatment specs from TEDS Output
Generate treatments
Delphi Method to create feature improvement specifications
Use method to generate feature improvement specs for participating organization
8

Pages 9 & 10

Evaluation Data from TEDS

- Ratings
  - Each Team (10)
  - Each scenario (8)
  - Each Cluster (12)
  - Conditional formatting to highlight good/bad
- Comments from raters
- Screenshots from raters

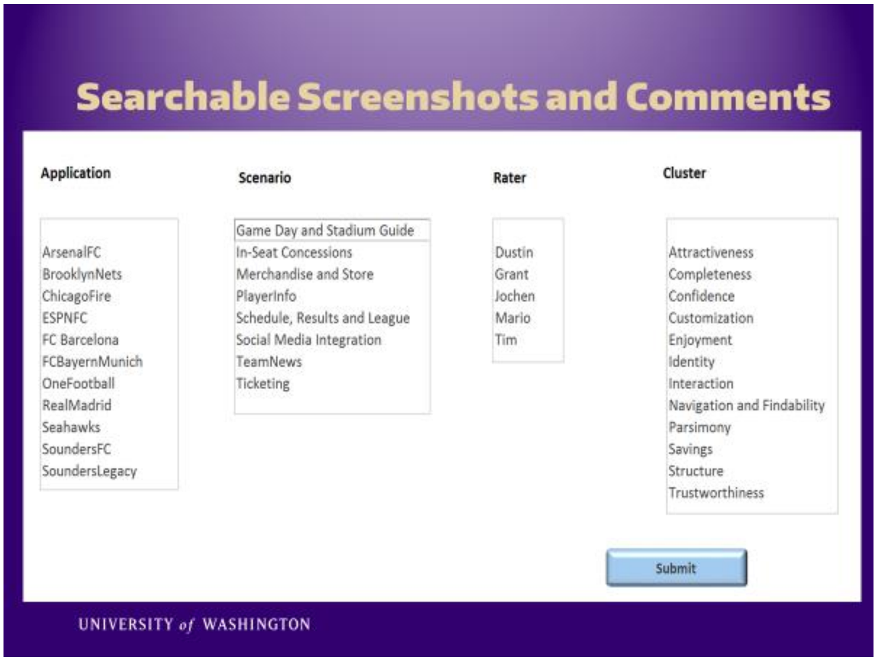
UNIVERSITY *of* WASHINGTON

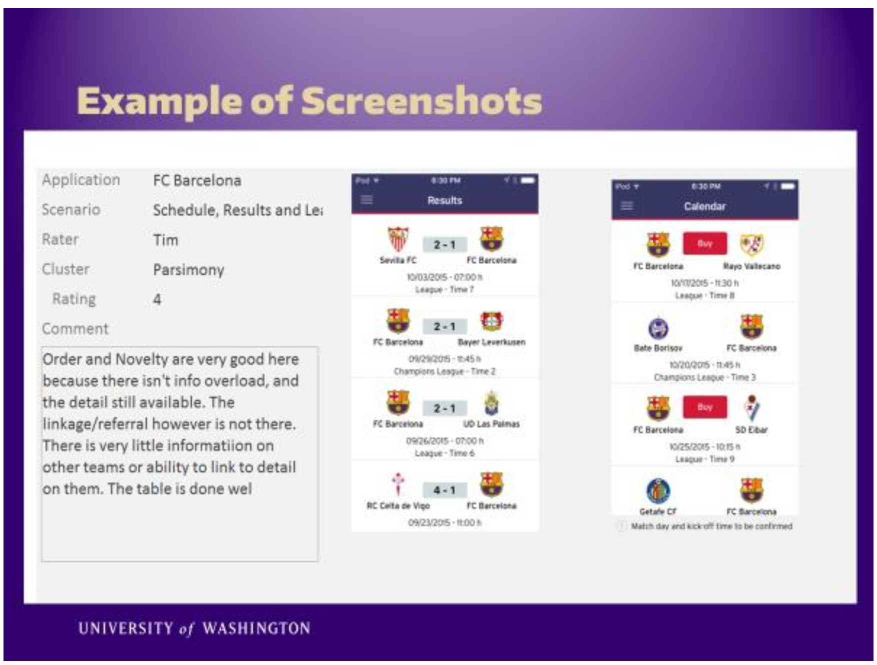
Overall average rating per Scenario

	Seahawks	FC Barcelona	Sounders FC Legacy	Real Madrid	Chicago Fire	FC Bayern Munich	OneFootball	ESPN FC	Brooklyn Nets	Arsenal	Sounders FC Overhaul	Average	Std Dev
PlayerInfo	3.45	3.58		3.79	2.79	3.04	3.64	2.17	3.36	3.13	3.35	3.23	0.48
TeamNews	3.69	3.64		3.30	2.81	2.92	3.75	3.13	3.30	3.03	3.36	3.29	0.33
Schedule, Results and League	3.83	4.10		3.38	2.89	2.98	3.95	3.19	3.72	3.25	3.25	3.45	0.42
Merchandise and Store	3.74	3.61		3.14	2.94	3.14			3.09	3.09	3.20	3.24	0.28
Ticketing	2.66	4.23		3.53	2.78				3.35	2.73		3.21	0.61
Social Media Integration	2.98	3.44		3.33	3.18	2.66	3.39	2.63	3.17	3.00	2.74	3.05	0.30
Game Day and Stadium Guide	3.19				2.58	2.66			3.48			2.98	0.43
In-Seat Concessions	3.15				2.06				3.52			2.91	0.76
Average Likert	3.34	3.77		3.41	2.75	2.90	3.68	2.78	3.37	3.04	3.18	3.22	0.36

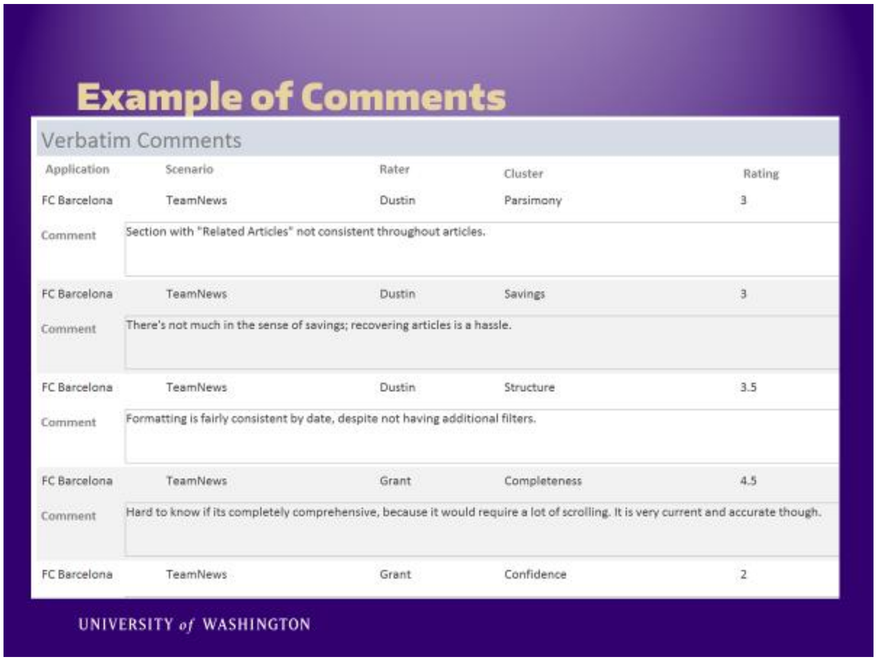
UNIVERSITY *of* WASHINGTON

Pages 11 & 12


Searchable Screenshots and Comments
Application
Scenario
Rater
Cluster
ArsenalFC
BrooklynNets
ChicagoFire
ESPNFC
FC Barcelona
FCBayernMunich
OneFootball
RealMadrid
Seahawks
SoundersFC
SoundersLegacy
Game Day and Stadium Guide
In-Seat Concessions
Merchandise and Store
PlayerInfo
Schedule, Results and League
Social Media Integration
TeamNews
Ticketing
Dustin
Grant
Jochen
Mario
Tim
Attractiveness
Completeness
Confidence
Customization
Enjoyment
Identity
Interaction
Navigation and Findability
Parsimony
Savings
Structure
Trustworthiness
Submit
UNIVERSITY of WASHINGTON


Example of Screenshots
Application FC Barcelona
Scenario Schedule, Results and Le:
Rater Tim
Cluster Parsimony
Rating 4
Comment
Order and Novelty are very good here because there isn't info overload, and the detail still available. The linkage/referral however is not there. There is very little informatiion on other teams or ability to link to detail on them. The table is done wel
Results
Calendar
UNIVERSITY of WASHINGTON

Page 13


Example of Comments
Verbatim Comments
Application Scenario Rater Cluster Rating
FC Barcelona TeamNews Dustin Parsimony 3
Comment Section with "Related Articles" not consistent throughout articles.
FC Barcelona TeamNews Dustin Savings 3
Comment There's not much in the sense of savings; recovering articles is a hassle.
FC Barcelona TeamNews Dustin Structure 3.5
Comment Formatting is fairly consistent by date, despite not having additional filters.
FC Barcelona TeamNews Grant Completeness 4.5
Comment Hard to know if its completely comprehensive, because it would require a lot of scrolling. It is very current and accurate though.
FC Barcelona TeamNews Grant Confidence 2
UNIVERSITY of WASHINGTON