Spring 5-2016

TOWARDS BUILDING AN INTELLIGENT INTEGRATED MULTI-MODE TIME DIARY SURVEY FRAMEWORK

Hariharan Arunachalam
University of Nebraska - Lincoln, hariharana87@gmail.com

TOWARDS BUILDING AN INTELLIGENT INTEGRATED MULTI-MODE TIME DIARY SURVEY FRAMEWORK

by

Hariharan Arunachalam

A THESIS

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfilment of Requirements

For the Degree of Master in Science

Major: Computer Science

Under the Supervision of Professor Leenkiat Soh

Lincoln, Nebraska

May, 2016

TOWARDS BUILDING AN INTELLIGENT INTEGRATED MULTI-MODE TIME DIARY SURVEY FRAMEWORK

Hariharan Arunachalam, M.S.

University of Nebraska, 2016

Advisor: Leenkiat Soh

Enabling true responses is an important characteristic in surveys; where the responses are free from bias and satisficing. In this thesis, we examine the current state of surveys, briefly touching upon questionnaire surveys, and then on time diary surveys (TDS). TDS are open-ended conversational surveys of a free-form nature with both, the interviewer and the respondent, playing a part in its progress and successful completion. With limited research available on how intelligent and assistive components can affect TDS respondents, we explore ways in which intelligent systems such as Computer Adaptive Testing, Intelligent Tutoring Systems, Recommender Systems, and Decision Support Systems can be leveraged for use in TDS. The motivation for this work is from realizing the opportunity that an enhanced web based instrument can offer the survey domain to unite the various facets of web based surveys to create an intelligent integrated multi-mode TDS framework. We envision the framework to provide all the advantages of web based surveys and interviewer assisted surveys. The two primary challenges are in determining what data is to be used by the system and how to interact with the user – specifically integrating the (1) Interviewer-assisted mode, and (2) Self-administered mode. Our proposed solution – the intelligent integrated multi-mode framework – is essentially the solution to a set of modeling problems and we propose two sets of overreaching mechanisms: (1) Knowledge Engineering Mechanisms (KEM), and (2) Interaction Mechanisms (IxM), where KEM serves the purpose of understanding what data can be created, used and stored while IxM deals with interacting with the user. We build and study a prototype instrument in the interviewer-assisted mode based on the framework. We are able to determine that the instrument improves the interview process as intended and increases the data quality of the response data and is able to assist the interviewer. We also observe that the framework's mechanisms contribute towards reducing interviewers' cognitive load, data entry times and interview time by predicting the next activity.

Acknowledgments

This thesis is part of a larger team project at the University of Nebraska – Lincoln (UNL), and thus would not have been possible without the opportunity to work closely with multiple students and faculty from both the Intelligent Agents and Multiagent Systems (IAMAS) group from the Computer Science and Engineering Department and from the Survey Research and Methodology (SRAM) group at UNL. I would like to thank Gregory Atkins, Dr. Adam Eck and Dr. Leenkiat Soh for the tremendous amount of their help in forming ideas and guiding me through this thesis through their advice and guidance. I would also like to thank Dr. Robert Belli, Dr. Alan McCutcheon and Ana Lucia Cordova-Cazar for taking the time to lend their expertise of the survey domain that helped shaped this thesis.

I would also like to thank my mother, Seethalakshmi, for being my rock and for her courage as we fought through one of the toughest times in our family. Words cannot express how much I owe her.

I also thank my Advisor, Dr. Leenkiat Soh, for his endless patience and support, tested through the duration of this thesis, without whom, I would have been none the wiser and this thesis, naught. Thank you sir for not giving up on me.

And finally, I thank all the energies – known and unknown, from all corners of this universe for all its blessings – in disguise and otherwise.

This material is based upon work supported by the National Science Foundation under Grant No. SES-1132015.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Introduction

1.1 Surveys and Issues

Surveys can be imagined to be human beings' way of attempting to quantitatively measure the perceptions of some population in society. By and large, surveys are seen by many researchers, developers and influential bodies as instrumental in having reinforcing effects and capable of providing a broader view or perspective at organizational and community levels. For example, governments and surveys share a ubiquitous relationship and it is believed that the outcome of several important surveys are responsible for government level attitude and policy changes. Indeed, governments have been known to use subsequent survey data to gauge the effects and implications of such changes. Hence it may be realized that surveys are often tools employed to perceive and visualize both demographic and/or temporal characteristics of the populations of interest. Its importance and the complexity of the field of survey conduction itself lays the first cornerstone for this thesis. It is empirical that surveys provide a sense of opinion of the targeted population and it may be expected that the targeted population may in return *expect* the opinions to make a difference (Page and Shapiro, 1983).

Notwithstanding the importance that surveys hold with the target population itself; another important factor notable is *enabling true responses*; the answers to survey questions must be the actual opinion of the individual, uncorrupted by any temporal or biasing effects induced by the conduction of the survey itself. This may include different forms of deception and socially desirable responding (SDR) (Paulhus, 2002).

Respondents of a survey must not feel that they simply must get through the survey and attempt to prioritize ease of completion over the *truth*. Thus surveys must promote self-disclosure and reduce deception (Hancock, 2007). Over the years, the way surveys have been conducted has undergone significant changes. It has transitioned, with continuing overlapping simultaneous steps, through different modes of conduction, from *face-to-face (F2F)* to paper-based, telephone-based and computer software-based (Conrad et al, 2007). Today, web-based surveys or Internet surveys are the latest models of survey delivery that is gathering momentum and favor with many survey methodologists, business interests and government bodies because of the ease with which it can be administered, collected and consolidated and for its better response rate (Cobanoglu et al, 2001). Software instruments for conducting the surveys have also evolved in how they are used and for what they are used for. This improvement in software instruments for conducting surveys however has not caught up with the improvements that the Computer Science field has to offer.

Systems such as Computer Assisted Telephone Interview (CATI) (Shanks, 1983) are built around a computer software that the *interviewer* would interact with and use to record the *respondent's (*who is being interviewed) responses. Online web surveys on the other hand deliver the surveys to the *respondents* directly over the Internet (Couper, 2000). In both cases, the responses are recorded through computer software and this has been the primary purpose played by the software during the survey process. This lays the next cornerstone for this thesis; the purpose of the software during the survey/interview. To better comprehend the objective of extending the purposes of the software, it may be worthwhile considering the track of the survey medium transition over the years. F2F is believed to be the golden standard for survey interviews by surveyors, since it creates a social presence that the respondent can actively interact with using visual, verbal and emotional actions. The disadvantage of the F2F approach is that the *social presence* can act as a deterrent when respondents are required to delve into personal and sensitive subjects such as sexuality, alcohol and drug usage (Currivan et al, 2004). However, Joinson (2001) showed how web and paper surveys have been instrumental in extracting truthful information without the associated *awkwardness* and reluctance of F2F. Thus while online surveys (surveys conducted with a software instrument) eliminates the advantages of F2F such as the interviewer's ability to detect whether the respondent understands the questions of the survey (by virtue of paralinguistic cues like pauses, intonation, speech disfluencies, gaze, posture and facial expressions (Graesser et al, 2008)) and personalize the interview as required, they can overcome the disadvantages of the F2F awkwardness.

The traditional *personal* touch lent to F2F interviews is thus unavailable in online surveys leading towards a generally more boring and dull perspective of surveys to respondents. Subsequently, the integration of F2F features into online surveys have been a topic of rising interest. It has been studied and observed that even a most minimal form of animation; as little as a line drawing animation, can invoke social behavior in respondents of non-survey tasks like personality rating (Reeves et al, 1996). This brings into question on ***how to define and understand what and how much assistance the software instrument can provide*** as part of the *personalization feature of F2F* since in the domain of surveys true response is significantly more important than speed or convenience. This does not however mean that we should completely disregard the advantages of online survey administration either.

Research in using visual assistance for surveys has made slow progress for questionnaire format surveys, and none is available for time use surveys. From the psychology point of view, Conrad (2015) offers the most progressed research on how using a virtual agent modeled to look human affects the respondents when answering questionnaire surveys. We examine this research in Chapter 2.

This places us within reach of the context of this thesis; which is an attempt to formulate a unifying solution framework for a specific survey system that would integrate the advantages of both F2F and online surveys whilst being within agreement of the principles of survey methodology. Current progress in this direction is little to none as the focus from the Computer Science point of view has been to perform the straightforward objectives of the survey and progress has been along the lines of how to enhance the instruments using technological features such as extended hardware and peripheral information gathering (such as GPS) without considering the effects of collecting and using such data in the context of surveys (Stopher et al, 2007). This vision is thus significant and this thesis aims to set a base track for future work to start from.

1.2 Time Diary Survey

Surveys are categorized and evaluated in a multitude of ways for theoretical and practical purposes. To help evaluate our proposed solution framework, we however, are specifically focused on one type of survey: Time Diary Survey (TDS). The American Time Use Survey (ATUS) is an example of a TDS. ATUS is a time diary type survey where the objective is to measure the amount of time American people spend doing various activities such as paid work, childcare, volunteering, and socializing. It provides nationally representative estimates of how, where, when and with whom Americans spend their time, and is the only federal survey providing data on the full range of nonmarket activities (from the website of the Bureau of Labor Statistics (BLS), United States Department of Labor at http://www.bls.gov/tus/overview.htm#1 which is the entity responsible for ATUS). Sponsored by the BLS, ATUS is conducted every year by the U.S Census department. Such nationally reaching time diary surveys are also conducted by other developed nations such as the United Kingdom, Germany, and Netherlands etc. ATUS is a CATI system and uses a software instrument that was internally built and is maintained by the Census department. A typical TDS's underlying purpose is to record the chronological sequence of events in a time-frame of respondent's life. Depending on the purpose of the TDS, the events may be day to day activities such as eating, drinking, working etc., or important events such as health based events, for example tobacco based events (starting, daily use, quitting etc.). The time-frame may also be short (a day) or long (months and years).

The ATUS in particular records all activities reported by the respondent over a 24-hour period from 4 am the day before the interview to 4 am of the interview day. It is conducted by a trained interviewer who uses a telephone to talk to the respondent and records the responses in the software instrument. This is a software assisted *interviewer* *based survey.* Interviewers are instructed to follow certain *scripts* and *probes* when conducting the interview to extract the relevant data, but are allowed sufficient leeway for interacting with the respondent.

This interaction is in terms of what form the questions are asked to the respondents and the interviewer's responses to the respondent's replies. We assume that the interviewer aims to keep the survey as short as possible while trying to gather all the required information accurately from the respondent. The interviewers also attempt to engage the respondent and keep the interview interesting to prevent the respondent from leaving the survey before completion. When a respondent leaves an interview before its completion, it is known as a *break-off*. The interviewer asks the required questions to the respondent and simultaneously fills out the respondents' responses (known as response data) in the instrument. The interviewer is thus expected to both maintain a conversation with the respondent while also interacting with the instrument to record the data. The current instrument used for ATUS is almost purely a data recording instrument.

1.3 Research Problems

Our research topic is thus the design and development of an intelligent integrated framework that is suitable to administer TDS under two modes: an *interviewer assisted* mode (IAM) and a *self-administered* mode (SAM). In the interviewer assisted mode, the system interacts with the interviewer who interacts with the respondent (directly or over the telephone). In the self-administered mode, the respondent directly interacts with the system.

The objective of the framework thus brings about two primary questions– (1) how to model the interview process and (2) how to interact with the user within the rules of the survey domain. It must be stated here that when the term *user* is used, it is assumed that it can be either the interviewer or the respondent depending on the context or both if used as an umbrella term. Any situation that warrants identifying a particular type of user would do so. The two questions thus raised are further reduced into their component problems and this work attempts to setup the path to realizing the framework and hence attempting to answer our questions.

The question of how to model the interview process is raised due to the nature of the problem that the framework is attempting to solve. TDS are essentially conversational surveys wherein either the interviewer or the respondent (or both) primarily control how the survey proceeds. For example, the interviewer may choose to ask the respondent to recollect from either 4 am the previous day (ATUS) or from another point of time that the respondent recollects. The respondent may also choose to start the conversation with recalling the activities or with general day to day conversations such as how they feel. The activities may be filled in order or depending on how the respondent recollects it. Though the intention of the survey is to extract the activities in chronological order, there are many different ways to accomplish this. The interviews are thus open-ended and of a free-form nature unlike conventional questionnaire based surveys. The successful completion of an interview may be verified by the presence of continuous and valid records for the entire duration as required by the survey (24 hours in the case of ATUS). However, other characteristics of the interview such as the speed, handling difficulties in recall and maintaining the respondent's motivation are not directly verifiable or even quantitatively or qualitatively assessable.

The other aspect of the problem of how to interact with the user is more open-ended. It cannot be assumed that the respondent also aims to complete the survey with complete and true responses. During the course of the interview, the respondent may have difficulties in recalling the activities they did or may be uncomfortable in recalling them. They may lose the motivation to continue with the interview if it is too long or boring from their point of view. The onus of keeping the respondent engaged thus currently rests with the interviewer who uses their expert interviewing knowledge to keep the interview on track as much as possible. Eliciting the required responses is the objective of the interviewer and he or she may employ conversational techniques and recall techniques to guide the respondent through the interview. Currently, the instrument used for administering ATUS is used by trained expert interviewers who can seamlessly conduct the interview over the phone with the respondent while entering the respondent's responses in the instrument. Understanding how the interviewer accomplishes this is not easily defined. In most instances, the interviewer may themselves be not aware of all the knowledge they possess or use during the course of the interview. The question to ask here is thus: how does an interviewer interact with the respondent and the instrument during the interview? This can be supplemented by another question: How should the instrument interact with the respondent? Given that the interviewers are trained extensively on how to use the instrument, can the instrument itself be leveraged to assist the respondent directly (without an interviewer) and thus reduce their cognitive load to the same level (or lower) as when they interact through the interviewer?

1.4 Motivation

The motivation for this work is derived from realizing the opportunity that an enhanced instrument can offer to the survey domain. Web based surveys can reach more respondents and are easily deployable. They provide faster response speed and have been shown to increase response rate and reduce the overall cost of conducting the survey (Cobanoglu, 2001). But training interviewers to keep up with an increasing respondent pool potentially increases the base cost of conducting the survey due to the time required to train interviewers and the subsequent running cost for each interview in terms of the time required and other related resources. This brings about the opportunity to unite the various facets of web based surveys to create an ***intelligent integrated multi-mode survey framework*** for delivering TDS. We envision the framework to provide all the advantages of web based surveys and interviewer assisted surveys so as to gain better foothold as a survey delivery method.

An intelligent framework is important since it would allow the survey instrument to partially take on the role as the interviewer and guide the respondent directly if required, thus eliminating the use of expensive interviewers. The framework must be multi-mode so that it can cater to both respondents (self-administered mode) and interviewers conducting surveys (interviewer-assisted mode). The integration of the two modes (IAM and SAM) allows for the instrument to be reused, thus reducing the time taken to develop instruments for each mode separately. Thus the intelligent integrated multi-mode survey framework would be able to scale and reach a wide range of audience, be easy to deploy and be usable by both interviewers and respondents without requiring to be extensively modified. The framework would further provide sufficient placeholders to extend it to power post processing of the data obtained through it. An intelligent framework can also provide personalization support, making it more appealing to respondents.

Another motivation is the potential to view the framework as a generalized solution to not just surveys, but other domains that would require similar interactions and structuring such as for hospital systems. This would effectively reduce the time required to formulate an alternate framework and subsequently enhance each other with their respective approaches to solve similar problems. Thus we can envision multiple scalable solutions from one single framework's underlying principles. Furthermore, once a solution is vetted and proved to work for the survey domain problem, it provides a strong ground for other similar solutions thus enabling them to be created and tested faster.

Since the intelligent integrated multi-mode framework is web based, it also provides for consistency when required and adaptability otherwise. For example, the instrument could change its representative form (the GUI) depending upon the device where it is accessed from, making it easier to use. Since the data would essentially be stored at a remote location, any instance of the instrument could use the data obtained from any number of previous instances to increase its overall effectiveness. Here by an instance of the instrument, we mean the copy of the instrument that would be used a user. A software instrument is more robust and can mitigate and recover from errors much faster than conventional means thereby reducing loss of data or respondents making it more appealing when considering the huge population samples such as is for ATUS.

1.5 Challenges

One of the biggest challenge to build the framework is understanding what data to use and how to use the data. ATUS defines a comprehensive list of activities (see Appendix 7.1) known as coded activities that each of the activity reported by the respondent must be categorized into. The process of converting the verbatim responses (word-to-word response given) of the respondent to its corresponding coded activity is known as coding. This is an intensive process and the resulting codes are not conversationally valid. This sets the challenge that the data in the coded form must be converted and used to power the framework that cannot use the coded activities as such. Also there are over 300 activities specified and a majority of them would hardly occur within a set of respondents – so how does one use this sparse data? While the absence of a particular coded activity in the data would prevent it from being used, it would nevertheless need to exist within the system. Furthermore, the system would not have any data to start off with if it were to use the data it generates to update itself. This is known as the cold-start problem. This is further complicated by the fact that there are many ways for the respondent to word their responses while only a fixed set of activities are recognized. How does the system figure out which coded activity the respondent's response corresponds to?

Another challenge arises out of how to interact with the user. While interviewers are considered to be fully motivated and hence are assumed to have no negative interactions with the instrument, the same cannot be said for the respondent. Constantly probing the respondent to check it an interaction is valid or not could potentially force the respondent to abandon the interview and break-off. Also, the process of conversing with the respondent by the interviewer is highly complex and may not be completely enumerated. How should the system react to respondent behavior? A simple rule based approach may be infeasible due to the many vastly different ways in which the respondents can behave. Given that an interviewer holds a conversation, only if the instrument itself is able to guide the respondent in a conversational manner when required will it truly achieve its optimal performance.

1.6 Proposed Solution Approach

Having a framework that works for both interviewer-assisted mode and self-administered mode in an integrated manner provides us with two major advantages:

1. Standardizes the data received from both the modes thus enabling the framework to use the data from interviewer-assisted mode to provide intelligent features to the self-administered mode.
2. Reduces the time required to adapt the system to the two modes of survey administration thus unifying the survey instrument rather than having to develop both separately; also reduces the time between converting elicited knowledge from data to operationalized knowledge in terms of software design and implementation.

Our proposed solution – the intelligent integrated multi-mode framework—is essentially the solution to a set of modeling problems. We model the data used for and within the framework, interviewer and respondent behaviors and break-off characteristics. This is accomplished by viewing the problems stated in Section 1.3 as the core focus. We propose two categories of overreaching mechanisms: Knowledge Engineering Mechanisms and Interaction Mechanisms to deal with the aforementioned problems. The Knowledge Engineering Mechanisms (KEM) are our solution to enable Knowledge Engineering (KE) within the system. KE deals with the processes involved in creating or transforming information into a form that can be used by a Knowledge-based system (KBS) (Studer et al. 1998). The various facets of this process includes everything from acquiring the knowledge (known as elicitation) to using the knowledge within the system. The various steps involved in KE are elicitation, analysis, construction, representation, validation, and maintenance (Ford et al. 1993). KEM thus serves the purpose of understanding what data we can create, use and store and subsequently how to use and maintain the data thus generated. The data thus generated with our KEM can be understood as the expertise of the domain in a form that is usable within our framework. Since the framework is an integrated one, our KEM pays special attention on how to separate the expertise required for interviewer-assisted mode and for the self-administered mode. We propose some ways to perform KEM by extending technologies used in other domains (similar and/or related) such as Recommendation Systems (RS), Case Based Reasoning (CBR) and Intelligent Tutoring Systems (ITS). We delve into the process of KE used within these domains and emerge with many KEM that are suitable for our integrated survey framework. We also examine and propose ways to understand how to use existing knowledge (or historical data) from the current version of ATUS to kick start the framework in its initial phase and thus provide a solution to encounter cold start issues. The components that make up the KEM of the integrated framework are classified into two different spheres based on their execution approach. These are (1) Online learning, where the system is live and in use and (2) Offline learning where the system is not in use. Offline learning may be conducted completed independent of where the system actually exists since it transforms the incoming information into the format required from it (expertise). This feature can be leveraged when the mechanism's execution might require significant computing power and time, without having to put those requirements on the live system. For example, supercomputers could be used to analyze the existing ATUS data (which runs close to a million records) and this processing can be done ahead of time thus relieving the system of requiring to have higher processing power adding to the scalability of the framework.

The Interaction Mechanisms (IxM) of the framework are those mechanisms (or components) that deal with the process of interacting with the user. IxM also maintains a separation between those mechanisms that involve interviewers and the ones that involve the respondents. This separation is important in the case of IxM because the interviewers and respondents are not equal in their commitment to complete the survey. This arises from the different motivating aspects for interviewers and respondents. While the framework could potentially require that the interviewer provide a constant stream of feedback while using the instrument, the same cannot be said for respondents using the instrument. Adding such a cognitive load on the respondent could potentially lead to break-off. Since the purpose of the integrated framework is to take the place of an interviewer in SAM, it performs different roles in IAM and SAM. This difference is thus accounted for by viewing the interviewer in the role of an expert user and the respondent as a novice user. The framework describes different mechanisms to interact with both the interviewer and respondent, the interviewer alone and the respondent alone.

Dividing the framework into KEM and IxM in no way separates them completely. Instead, by adding this division we simply create two areas of concerns that need to work synergistically, but can solve their respective problems independently thus allowing for a high level of modularization during implementation. This can further increase the efficiency and scalability of the system.

1.7 Contributions

The primary contribution of our work is in paving the way to make an intelligent multi-mode survey framework that is capable of conducting time diary surveys under the two modes: interviewer assisted mode (IAM) and self-administered mode (SAM). With this endeavor, we make forays into three primary fields: (1) Computer Science, (2) Survey Research and Methodology, and (3) Survey Informatics. In the field of Computer Science, we contribute to the areas of Intelligent User Interfaces (IUI), and Recommender Systems specifically with respect to restrictive environments such as time diary surveys that are characteristic of bias, restricted feedback and knowledge elicitation. In terms of survey research and methodology, we primarily contribute towards a multi-mode time diary survey instrument with our prototype instrument in the interviewer-assisted mode. Our contribution extends towards computer assisted telephone interview (CATI) systems and adaptive designs for time diary surveys. Our contribution towards the domain of survey informatics is the prototype framework implementation that enables the use of tracked paradata from interviews to improve how the system interacts with the users. We briefly enumerate our contributions below, and expound on these in Chapter 6:

1. Computer Science
    a. Use of Intelligent User Interfaces (IUI) and Recommender Systems (RS) in restrictive environments (bias, restricted feedback) with knowledge elicitation,
    b. Integrated framework for multi-mode time diary survey administration,
    c. Prototype framework instrument based on our framework in IAM
    d. Generated response data and paradata for future work in SAM.
2. Survey research and methodology
    a. Instrument prototype demonstrates assistive CATI time diary system,
    b. Adaptive design for surveys,
    c. Designed and implemented paradata logging and tracking
    d. Using historical data for eliciting domain knowledge
3. Survey Informatics
    a. Integrated framework that enables use of paradata to improve interviews.

1.8 Overview of Thesis

In the next chapter, we describe the background and related work for time diary surveys and review the literature regarding Computer Science technologies and methods that are applicable to time diary surveys. In Chapter 3, we delve into the details of the fundamental research problem and describe the methodology by which we build our intelligent integrated multi-mode time diary survey framework. Then, Chapter 4 gives the technical details of the prototype implementation of our framework in IAM mode. Chapter 5 presents the results and the analysis of the experimental studies performed using the prototype implementation. Chapter 6 then gives the conclusions about our work as well as ideas and directions for future work. Chapter 7 lists the various accessory items in the form of an appendix.

Chapter 2: Background and Related Work

2.1 Introduction

In this chapter we delve into the background work and existing technologies that play a fundamental role in shaping our work. In Section 2.2, we begin by first examining the approach taken by survey research to leverage the techniques of Computer Science. Then, in Section 2.3, we extend our examination towards time diary surveys and the challenges associated with computerizing them. We pay special attention on the multi-mode aspect of this. As a reminder, by multi-mode we mean the ability of a singular instrument (or framework) to address both self-administration of the survey by the respondent directly and interviewer-assisted administration by an interviewer. After addressing the survey research side of our framework, we move on to the technologies in the field of Computer Science that cater or has potential application to the survey domain in Section 2.4.

By understanding the principles and background behind these technologies and associated techniques, we can fully appreciate the need for an intelligent integrated multi-mode survey framework for time diary surveys (TDS) and the advantages that Computer Science can offer to create a more robust and usable framework for administering surveys.

2.2 Surveys and Instrument Design

As described in Chapter 1, surveys are an important tool used by many fields to collect and analyze opinions and information regarding a target population. The intricacies of defining the objective, purposes and characteristics of a survey primarily falls under the survey methodology domain. This meant that survey methodologists put together the required specifications of the survey such as the target population (e.g. nationalities within a country, specific professions etc.), the purpose of the survey (e.g. political opinion, medical history, genealogy, time diaries etc.), and how it is to be administered (e.g. face to face, paper based etc.) so as to get data as good as possible. While they can be considered to be the experts for defining the format of how the questions of a survey should be worded and formatted, a new visage of survey administration has emerged with the advancement to web based and online surveys. To help better understand the differentiation between interviewer administration and the online development of conventional surveys and the concept of TDS, we first briefly examine the questionnaire format survey (the conventional survey) before going into time diary surveys. This allows us to understand the unique differences between questionnaire surveys and time diary surveys. Once we examine these differences, we examine TDS more closely and look at the current efforts in improving time diary surveys together with existing instruments that are used to administer time diary surveys.

2.2.1 Questionnaire format survey

While originally an interviewer would serve as a medium between a respondent and the subsequent media of recording (paper or software instrument), the extent to which web-based survey administration has expanded the reach of the survey makes it harder and harder to employ such intermediaries to assist the respondent while making the prospect of delivering the surveys directly to the respondent more appealing (Andrews, Nonnecke & Preece, 2003). Furthermore, the type of the survey also influences the cognitive load on both the respondent and the interviewer during the process of a survey interview. For example, in a questionnaire format survey, the respondent is presented with a set of questions that can be answered by either picking from a pre-defined list of answers (or options) or wording the answer in free-form as the respondent's response. They may or may not contain skip patterns (depending on a specific question's response another question may become available or become unavailable), may or may not be mandatory (the respondent is free to not answer a question), or require a particular order in which it must be answered (Litwin, 1995).

When questionnaire format surveys are delivered via online web instruments (or simply survey pages), the design of the instrument usually follows the corresponding paper format. Research in this area is however fast paced and is exploring how the media used by the respondent to access the survey such as whether it's a simple mobile phone, or a smart phone, or a tablet or a personal computer can affect the format of the survey in a dynamic way. The web-based questionnaire delivery method provides advantages such as versioning, delivery control, recording and even post processing analysis. As such designing an instrument for the administering the questionnaire directly to the respondent is rather trivial once the specification is known (Dillman & Bowker, 2001).

As mentioned in Chapter 1, Conrad (2015) presents us with the most progressed research on how a virtual agent affects respondent behavior in questionnaire surveys. They use a virtual interviewer modeled on a human face to take on the role of the interviewer and tested two modes: (1) High versus low facial animation, (2) High-dialog-capability versus low-dialog-capability. Facial animation varied the amount of facial expressions for the interviewer between high and low, while dialog-capability varied how many dialogs the virtual interviewer offered the respondent. In their work, participating respondents interact verbally and visually with the virtual interviewer, which is “wizarded”. This means that the intelligence of the agent for responding to the respondent’s visual and verbal responses was controlled by a hidden researcher – unknown to the respondent until the survey is completed. Thus it must be noted here that there is no active intelligence to the virtual interviewer – the research focuses on how a virtual interviewer would affect the responses and clarification behavior of the respondent. They report that, while the respondents provided more true responses (based on a fictional scenario to keep track of the true response) to a virtual interviewer that had high facial animation, respondents seem to not be affected by how they use the virtual interviewer to provide clarifications for the questions asked. The authors were unable to determine statistically significant evidence to support their hypotheses that respondents would engage more with high facial animation and high-dialog-capability virtual interviewers. They however, were able to observe suggestive and self-reported evidence that respondents preferred to interact relatively more with low facial animation and high-dialog-capability virtual interviewers. This work provides us with a little more understanding of how using animated virtual interviewers could potentially affect the respondents. While this research was based on questionnaire surveys, when we look at time use surveys, which are open-ended and free-form, it is more essential to consider how the virtual interviewer would assist the respondent (like a human interviewer would do). Since there is no substantive evidence that a “wizarded” virtual interviewer can engage and interact significantly better with a respondent, a step back would be necessary to understand how an intelligent virtual interviewer would be able to assist and engage with respondents in self-administered time diary surveys.

2.2.2 Time diary format survey

Unlike questionnaire format surveys however, time diary format surveys are intended to elicit and record the respondent’s time use data. Time use data is the chronologically ordered list of activities (and their context information) performed by a respondent during a particular time period. Time diary surveys (TDS) are generally conducted to record the respondent’s self-reported responses since this information is unavailable by conventional means of observation. TDS may be administered in a paper-based format, where the respondent fills out the survey form with the activities they performed by recollecting it (Horrigan, Michael & Herz, 2004). Just as questionnaire format surveys advanced with the introduction of computers, TDS has also moved forward in the same direction (Wright, 2005). However, the inherent complexity of time diaries has prevented it from advancing at the same pace. These complexities primarily involve the lack of structure in how time diary surveys are filled out, the increased cognitive load required to fill out time diary surveys and a lack of motivation for respondents to sit through time use surveys (Bolger et. al, 2003). Software instruments used to administer TDS are usually complex and require a significant learning curve and thus TDS are primarily administered using a trained interviewer who acts as the intermediary between the respondent and the instrument. While research exists on computerizing TDS, work done in exploring how TDS can be administered directly to the respondent via the web is being studied primarily from the survey point of view with respect to its issues and expected data quality. (Crosbie, 2006).

2.2.2.1 **Event History Calendars**

Event History Calendar (EHC) is a closely related type of survey to TDS in that they are designed to capture autobiographical information from a subject and place it on a grid where one dimension is time (Kite, 2007). Similar to TDS, EHC also requires respondents to recall events from their past. Thus, by examining EHC and their computerization and automation efforts, we can develop an understanding of the characteristics that would affect the design of a TDS framework.

2.2.2.2 **Previous efforts in EHC**

The work by Kite (2007) is a significantly advanced step towards an automated EHC framework that aims to substitute an interviewer with an intelligent software component. The approach used in this work leverages an adaptive conversational case retrieval system to replicate the conversation process of an interview between a respondent and interviewer in a self-administered setup. The intelligent interview system designed in his work takes upon the tasks of modeling the domain knowledge and of modeling the interviewer. Similar to the problems faced while developing an intelligent TDS framework, the automation of EHC faces human-computer interaction, knowledge modeling and user modeling (interviewer) problems. The data involved in EHC is also temporal, unstructured and subject to the respondent’s recollection ability. While in essence both our and Kite’s work focus on computerizing an interview assistant viewing it as a modeling problem, there are stronger differences in how this task is achieved and the overall objectives. His framework comprises of a knowledge engineering component for using and maintaining the domain knowledge and a phased implementation of an intelligent assistant using a modified Case Based Reasoning (CBR) system called Conversation Cased Based Reasoning (CCBR), while our framework focuses on Knowledge Engineering Mechanisms for modeling the interview, interviewer and respondent characteristics and Interaction Mechanisms to deliver the knowledge gained using the former. Thus both the frameworks effectively have two synergistic components that work in tandem. Our framework however takes a broader view of the problem and thus views the interview as a process with two distinct modes (multi-mode) – the interviewer-assisted mode and the self-administered mode, while his framework approaches this with a more detailed focus on the self-administration mode. Because of this distinction, our framework pays special attention to ‘who’ uses the system.

Kite’s framework’s knowledge engineering component performs knowledge acquisition using pattern recognition and data mining using an apprenticeship method, where it tracks and learns (creates cases by observing patterns) an interviewer using the instrument. Our Knowledge Engineering Mechanisms focuses more towards data mining from historical data and observed data in a multi-mode setup. Thus the apprenticeship method of learning is a subset within our Knowledge Engineering Mechanism. Our domain knowledge thus is the interview process itself rather than memory recall processes. Thus while the frameworks show a difference on how the mined data is being used, essentially both the frameworks take very similar approaches by using the paradata attained through methods with different objectives.

Another important aspect is the availability of verification methods in EHC which is absent in TDS. Since EHC focuses on landmark events, there exists rules such as, if a respondent reports being pregnant then it must end in child-birth, which can be checked for violations, thus creating space for truth-checks. In TDS however, such rules for truth-checks are hardly available and are broader. For example, a change of location between two activities without a traveling activity between them is such a violation. However, the respondent could have reported it using implicit wordings such as ‘I did A, and then I went over to X to do B’ making it a recording issue rather than a recall issue. The inability to verify the truth of the data reported in TDS makes it harder to create rules and generalized patterns.

Kite’s framework attempts to replicate the interviewer reasoning while eliciting information from a respondent while our framework attempts to provide assistance to the user (respondent and interviewer) for data entry, usage guidance while attempting to keep the respondent engaged and thus result in the elicitation and recording of the information.

His framework views the interview process in a (Question, Response) format while we view the interview process as a set of interactions between the user and the instrument. Another significantly distinct aspect is the focus of our framework to provide the means to handle noisy and erroneous data as an interviewer would do during the interview.

Thus while both the frameworks undeniably are attempting to solve the very similar problem of computerizing an interview process for information elicitation, they differ in the method of approaching this problem and in the ways it takes to provide the solution under the two similar, but not same environments of EHC and TDS. Both the frameworks attempt to reduce the cognitive burden on respondents in a self-administered setup, but Kite's framework does not keep that as an objective when an interviewer is the user which ours does. Thus Kite's work provides insight into how a computer-human interaction problem similar to TDS can be computerized and provide a basic understanding of how to replace a human-human interaction during information elicitation.

2.2.2.3 **Current efforts in Time Diary Surveys**

Research in computerizing TDS has been limited to primarily converting the paper equivalent of it on to a software application. The American Time Use Survey (ATUS) is a prime candidate for examination of the background in TDS since its inception was in the paper based format and it has evolved over the past two decades into a Computer Assisted Telephone Interview (CATI) format. Following various rounds of testing and field studies, they reported that an enhanced instrument that included probes that asked respondents if they stopped an activity to do another increased the data quality

(Forsythe, 1997) and later that due to concern about respondent burden, and the complexity involved in programming the computer software they would not attempt to collect secondary activities using the instrument (Horrigan & Herz, 2004). The instrument used for ATUS has undergone cycles of revision, but it is of our opinion that it has failed to fully leverage the advantages offered by the cutting edge technologies in the fields of machine learning, information filtering and human-computer interaction. The instrument still remains primarily as a tool to assist the interviewer in recording data and collating interviews. Section 2.3 examines the instrument used for administering ATUS and describes the functions of the instrument. Section 2.4 then examines two other significant related works in the area of computerizing time diary surveys. These examinations will further strengthen our motivation for working towards an intelligent integrated multi-mode time diary survey instrument.

Time diary surveys are thus characterized by the difficulties faced in helping respondents understand the process of completion, the way the instrument used interacts with the user (interviewer or respondent) and by the general rules of surveys that require a consistent, non-biasing approach to completing them.

The time use surveys we examine in the following subsections are characterized by the way they approach time diary surveys from the point of view of the survey domain. This delegates the implementation to the Computer Science field rather than approaching it from the point of Computer Science, wherein it could offer solutions to the problems faced in implementing time diary surveys.

Time use surveys can be characterized by the amount of information required for it to be considered as complete responses in the eyes of survey methodologists and the procedure through which this is extracted (Stinson, 1999). The respondent's responses for activities are not expected to be in chronological order; forcing this, for either reporting or recording, *increases* the cognitive load on the respondent or the interviewer respectively. Furthermore, activities require adequate context information—

who was with the respondent when they performed the activity; *where* was the respondent when they performed the activity. This context information is used by researchers to categorize activities accordingly. For example, 'eating' may be a 'work-related' activity is performed at the respondent's workplace or if the respondent was with co-workers (Stinson, 1999). When time use surveys are conducted by interviewers, they assist and guide the respondent in recalling their activities—they may do so sequentially or by backtracking or in the order that the respondent reports in. Since the respondent already faces the cognitive task of recollecting the activities, it may be unwise for interviewers to constantly ask for additional information that could detract the respondent from their task. This rationale leads to an environment of restricted feedback, wherein the respondent and/or interviewer may not be able to provide immediate feedback about the interview or the processes related to it. Thus we can see that time use surveys (1) are more open-ended, (2) requiring sufficient content information to be considered as complete responses while (3) limiting how much feedback can be obtained from the respondent.

2.3 American Time Use Survey (ATUS) Instrument

The instrument used to administer ATUS is prima facie intended to be part of a CATI system. It is a Graphical User Interface (GUI) based software that interviewers are trained to use and consists of the different screens required to manage respondent information, roster information (respondent’s household members), the time diary information and some demographic information. It is not web-based and the entire application must be downloaded to the user’s computer to be used. Figure 1 shows the user interface for the 2010 ATUS instrument, where the interviewer would record the activities and their context information reported by the respondent during the interview.

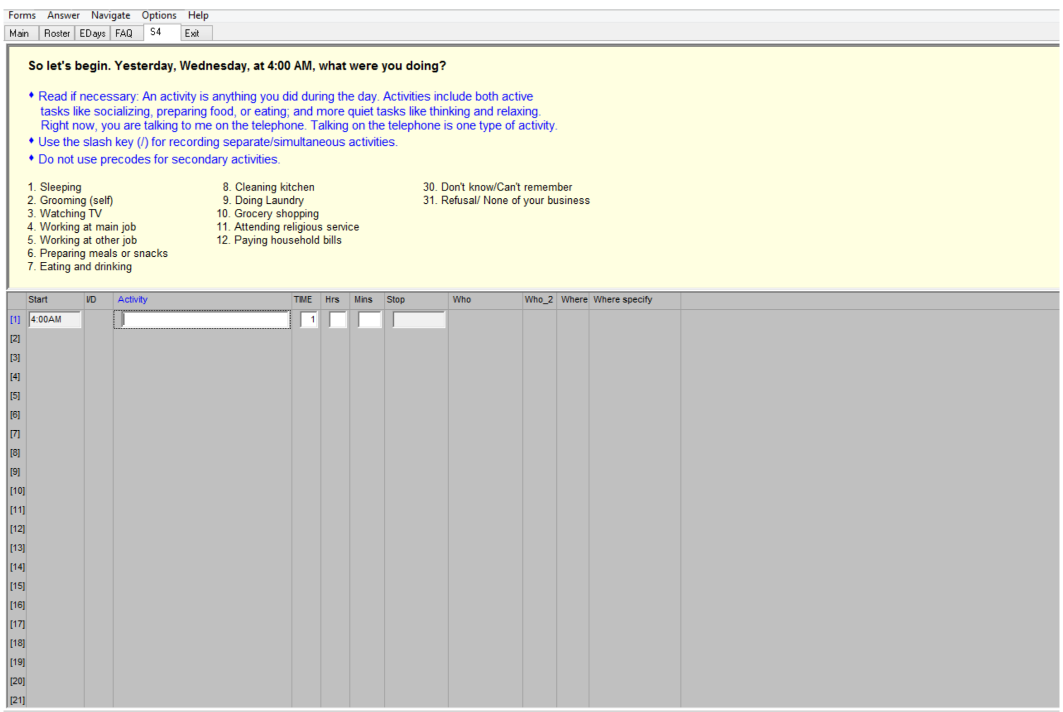

Figure 1 Current ATUS Instrument's Activity Recording Screen (2010)

As seen in Figure 1, the interviewers would enter the information of each activity in a list format. An information frame at the top half of the instrument provides the interviewer with a standard text that they can use to talk to the respondent. This text provides the interviewer with some general guidelines for both using the instrument and about the interview rules themselves. The instrument provides basic validation features such as range checks, duration validity checks, and activity coding checks. In addition to these validations, it also provides the interviewer with probes that pop up when certain conditions are encountered such as if an activity other than working or sleeping has a duration equal to or more than 3 hours (Figure 2). The interviewers are also trained and provided with the set of probing rules that the instrument provides (see Appendix 7.2).

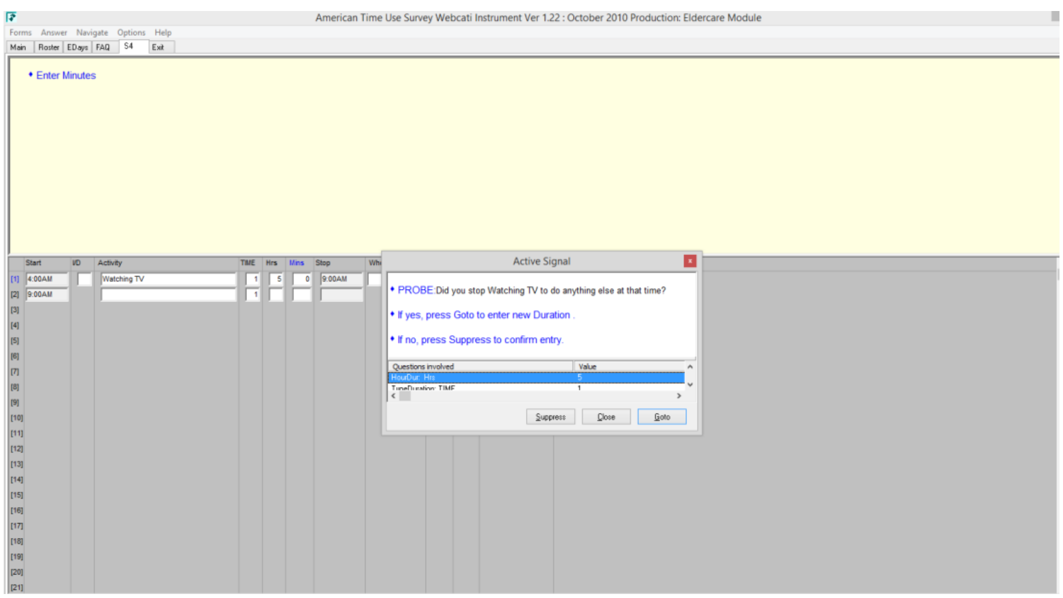

Figure 2 Current ATUS instrument's long activity duration probe

While the instrument provides sufficient functionality for a trained interviewer to use the instrument, it nonetheless requires a significant learning effort if it were to be used by a respondent directly for self-administration. Certain features such as the **Time field** accepts either a 1 or 2 where a value of 1 means that the end time is specified by providing the duration and a value of 2 means that the end time is specified by providing the time itself, are not intuitively designed to be understood at a glance. While reducing respondent burden is of significant importance in the field of survey administration, the current ATUS instrument does not provide any confidence for it. Furthermore, the activities are listed in a top-down format. While this makes it suitable to be read out by an interviewer during an interview, it does not provide a way to visualize the respondent's day in an easy manner. Thus it can be stated with some confidence that the instrument is primarily meant to be used by trained interviewers under a CATI setup. It makes no use of the data collected to improve itself nor does it observe the interviewer for understanding how the interview process works. Thus it is effectively a dumb instrument intended to perform the role of a data recording tool albeit with certain enhancements to make it easier for interviewers and a far cry from being able to be used for self-administration.

2.4 Other Time Use Survey Instruments

As mentioned in Chapter 1, time use surveys are conducted by many developed and developing nations to collect information about how people spend their time. While this has prompted development and research on refining the process of collecting and using the time diary data, the administration instrument itself has not been a primary focus mostly. In this section, we will examine two of the works that do lay some focus on the instrument design while considering the time diary surveys bigger objectives.

2.4.1 Harmonised European Time Use Survey

The Harmonised European Time Use Survey (HETUS) is a paper-and-pencil based time use survey administered in the European Union similar in concept to ATUS. Unlike ATUS however, HETUS is not a CATI system and respondents are provided with time diary sheets to fill out their daily activities. These sheets are then collected, coded, cleaned and digitized manually. This brings about a longer turnover time from the start of the survey to the final data publication. Also, due to its pen-and-pencil based approach, clarifications cannot be asked of the respondent's regarding the responses. Furthermore, since the fieldwork, coding, cleaning and digitizing is performed manually by trained personnel, it adds to the base cost of administering the survey. The focus of HETUS is primarily to perform data collection in a large demographic region (Europe) and currently does not focus on computerizing the process. However, one of the stated aims of HETUS is to create an automated intelligent time diary survey instrument - update on the progress of this aim was not available. Since the current efforts in HETUS are not aligned with our eventual goal of a self-administered online time use survey instrument, we do not delve into a detailed comparison between the two.

2.4.2 Modular Online Time Use Survey

The Modular Online Time Use Survey (MOTUS) is a full-scale implementation of a TDS system that attempts to create a more online web-based approach to designing, managing and administering time use surveys. The primary challenge addressed by MOTUS is to translate the typical paper-and-pencil time diaries to an online method without losing the strengths of the paper-and-pencil approach of not requiring expensive interviewer costs, with additional features to enrich the data and with automated processes to reduce personal and processing costs (Minnen et al, 2014). Thus it is essentially the first survey instrument implementation to truly embrace the embodiment of 21st century web technology. The first field-testing of the instrument was done in 2013 and the results published later in 2014. It envisions truly advancing the way time diary surveys are conducted by leveraging the reachability and large-scale administration capability of the Internet. Designed and developed the Research Group TOR of the Sociology Department of the Vrije Universiteit Brussel, it provides a complete suite of features for administering time use surveys such as (from the MOTUS official website):

- Direct Data Storage (DDS)

  Data inputted by the respondents are stored directly on the server and are thus immediately available.

- Respondent Management System (RMS)

  Provides the ability to import lists of respondents, manage them (assign usernames and passwords, change password etc.) and assign respondents to surveys and send out mass communications to the respondents.

- Respondent Tracking System (RTS)

  Provides the ability to monitor respondents while they use the time use survey recording paradata information like logging times, page load times, field entry and update times and the progress of the respondents. It also provides the ability to export progress reports and response rates for different elements of the time use survey.

- Customized Survey System (CSS)

  Provides the ability to completely customize the survey with respect to pre-interview questionnaires, post-interview questionnaires, and virtually all elements of the time use survey such as the activities hierarchy, skip patterns, contextual information of activities etc.

A screenshot of MOTUS' online activity entry page is shown in Figure 3. Each component of the activity information is separated into tabs (When?, What?, Where?, Whom?) at the top of the data entry area with a listed view of the activities on the right side along the border. The activity information can be entered using a multi-level combo box selection control or manually entered using a search facility. Activity context information (Where & Whom) provides a list of options to select from (e.g. 'Where' has home, school, other people's home etc.).

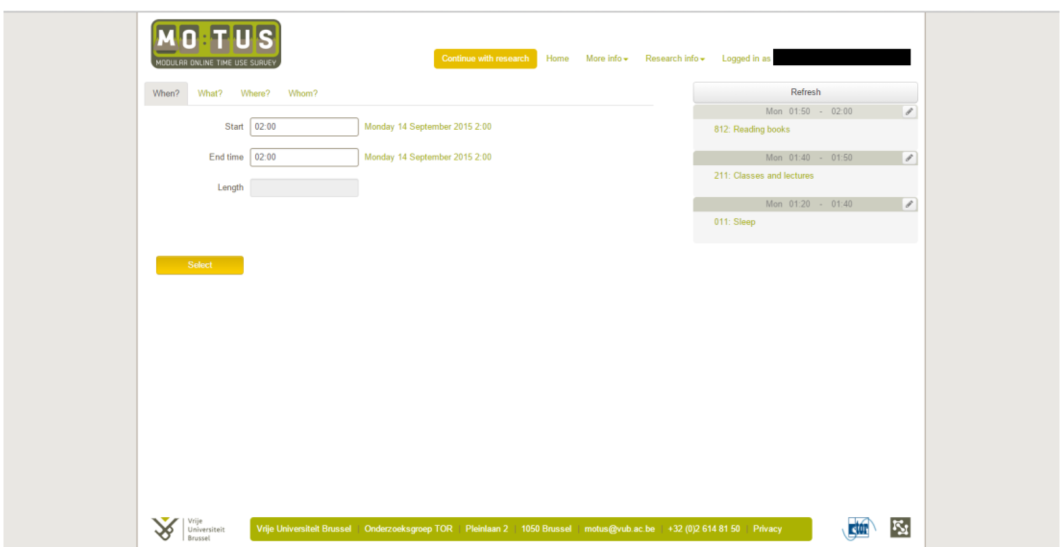

Figure 3 Screenshot of MOTUS activity page

The instrument comes built-in with both hard and soft warnings, where hard warnings must be handled before the activity can be saved while soft warnings can be ignored. An example of a hard warning is a missing end time, while an example of a soft warning is when an activity has more than 20 hours of duration. The system on examination has some advantages and issues.

First, the interfaces are clean and adapt to different screen sizes by following a fixed width design pattern. The design allows most Internet users to figure out what kind of information is being asked and how to provide them. A 3-tier hierarchical drop down control allows activity selection in a highly efficient manner. The type and search feature alternative to the drop down control is clean and provides sufficient autocomplete support. Provides displacement warning (when the location between two activities change without a traveling activity). Assumably the ability to completely manage the survey could be of significant use to survey designers, however we were unable to access this feature and lack the required qualifications to evaluate it.

The instrument, in our opinion, faces some issues both in its design principles when targeting the general internet savvy population, and its usability when targeting respondents. The activities need to be entered in a highly sequential manner making it susceptible to be boring. The visual representation of the chronologically sorted activities as a top-down list feels a bit dull. Hard errors are not indicated during the process of data entry causing revisits after submission. Each time an activity is saved, there is a brief period of non-response that could potentially affect user's interest. Delays such as this on the web are usually server induced and fall under the general term of *lag*. There is no progress indicator displaying how much more data is needed or the range of data needed. Furthermore, the unavailability of a confirmation window when closing before completing the survey could allow accidental closing. Thus, in the interactions aspect, the instrument may not feel engaging to the respondent. While it provides certain intuitivism to its usage, it lacks an overall flow structure aimed at assisting the respondents or in helping them complete the survey process. These disadvantages discourage respondents from completing the survey successfully.

The pilot-study (Minnen et al, 2014) demonstrated that their modules (read: “different and additional context information”) did not result in different respondent tendencies with regard to participation in MOTUS. What this means is that asking a few more questions for certain activities did not induce negative respondent learning (where the respondent would actively avoid providing those activities). They view the absence of an interviewer as an insurmountable obstacle to improve their response and participation rates. They also propose many changes to counter the different issues mentioned above in their future work.

MOTUS relates to our framework in the way it attempts to bring time diary surveys online and in targeting self-administration by respondents. However, our framework attempts to keep the instrument as a single screen interface – thus preventing users from having to encounter constant page loads. Furthermore, our work focuses on how to assist the user (interviewers and respondents) so as to reduce the cognitive load exercised during the interview; MOTUS is aimed at respondents alone and attempts to follow a questionnaire survey type flow in an attempt to make it easier for the respondents to complete the survey. Thus, while our work and MOTUS share some common objectives, differences exist in how the problem of administering time diary surveys is approached.

2.5 The missing link

On examining the different time use survey instruments currently used, we can come to one converging conclusion – the interviewer is an important aspect while conducting time diary surveys. Whether they ease the respondent's cognitive burden or raise the trust of the system or build a rapport with the respondent, they essentially bring in the advantages of Face-to-Face (F2F) (Chapter 1) to the interview one way or another. While this seems insurmountable from a survey standpoint, when we examine the issue from the Computer Science point of view, we realize that there are many technologies that are currently attempting to solve the very same problem in part or full in various other domains. Thus Section 2.6 is well placed to detail our examination of the different Computer Science technologies that can be leveraged for the purpose of incorporating intelligence into a human-computer interaction environment.

2.6 Current Computer Science technologies

2.6.1 Computer Adaptive Testing

The Computer Adaptive Testing (CAT) system is the more powerful successor to a series of successful applications of adaptive testing (Linacre, 2000). The objective of a CAT system is to determine within a margin of accuracy, the ability or skill value of a test taker by challenging them with pre-ranked questions on a difficulty scale. Depending on which variation/adaption of CAT is used, a transformation scale is selected that maps the difficulty of a question against the ability of the test taker when it is solved correctly. The process starts with the system choosing an arbitrary base point (average) difficulty question to the test-taker. If the test-taker gets the answer right, a higher difficulty question is asked, else, a lower or similar level question is asked (depending on if the system is gauging the ability or attempting to converge). This process repeats until it converges to a point where the test taker has a 50% chance of success/failure or a 100% chance of failure (depending on the model). Other exit conditions for the system include time limits and/or a preset number of questions.

The primary focus when examining the CAT system is on understanding the parameter of 'difficulty' – which may be pre-coded by the testing authority while generating the questions or determined by the system during a learning phase (research/test section) by analyzing the maximum difficulty level at which test-takers last succeeds at solving it or the minimum difficulty level that guarantees failure (Linacre, 2000).

Thus CAT systems are essentially aimed at modelling the student's ability against an arbitrary difficulty-ability scale (dichotomous Rasch model). Different systems/authorities adopt different types of scales and testing sequences depending on the method used such as the basic procedure (Binet, 1905), or the Flexilevel testing procedure (Lord's, 1980) and its variants such as the Step Procedure (Henning's, 1987) or the Testlets (Sheenhan's, 1990).

CAT systems mostly deal with dichotomous items, where every item has a difficulty expressed as a linear measure along the latent variable of the construct. The latent variable of construct is essentially the range of ability that is testable by the given set of questions. CAT systems have also been modified to work with polytomous items, but this is achieved by essentially breaking down the question to follow a pattern similar to that of the dichotomous items with partial credits. Furthermore, CAT systems must particularly or rather, mostly work along a one-dimensional variable which in most cases is the difficulty level of a question versus the ability of a student. The scale provides the correlation among them. Multi-dimensionality is known to confound the CAT process since it brings about ambiguity about what is the ‘correct’ and ‘incorrect’ answer. For example, if the dimensions being measured were mathematical ability and literacy, and a particular numerical question had a certain difficulty level in both dimensions, which dimension should be considered the reason for an incorrect answer-low literacy or low numeracy? Did the student fail to understand the question (low literacy) and hence fail to answer or did they understand the question but fail to apply the corresponding correct mathematical solution (low numeracy)? CAT systems view such multi-dimensionality as two uni-dimensional tests intertwined, and separate the test in such a way that for one that measures the numeracy ability, a basic literacy level is assumed and the questions are framed within those expected limits.

As discussed above, it is evident that a fundamental requirement for employing the CAT in a system be that the system have a uni-dimensionally observable variable. The integrated framework in its essence will have (i) multiple variables for observation such as interest, break-off probability, motivation and effort, (ii) users who are not as motivated. Test-takers of the CAT system are motivated to take the test for whatever reasons the CAT test is being administered, which is not the case with the surveys since it's more an obligation than a requirement and (iii) the system itself is motivated in testing the users to determine their ability. Thus one of the biggest driving factors for the CAT system, that both the participants are motivated is unavailable for exploitation in the integrated framework for survey system. That is, survey respondents are not all motivated to be truthful nor even complete the survey in one go, while the survey system has to work to keep the respondents engaged. Also, unlike the CAT system, the integrated framework cannot easily reduce the multi-dimensionality without considerably increasing the demands on the respondent, which is not an option and would destroy the survey altogether.

But nevertheless, we can draw some comparisons between the integrated framework and the CAT system. Both systems are measuring some variable of the user and mapping them to an internal scale. This modelling of the user, is a key component that is extended and adopted from the CAT system, onto the integrated framework.

2.6.2 Intelligent Tutoring Systems

"Intelligent Tutoring Systems (ITS) appeared during the 1970s were driven by the success of knowledge-based systems and expert systems" (Ramos et al, 2009). They are intended to be able to deliver subject knowledge to train students/professionals and verify the results of the training without involving human instructors. It was responsible for bringing about many ideas like using computational models of domains and intelligent reasoning and explanations. They are excellent examples of practical implementations of artificial intelligence, natural language, machine learning, planning, multi-agent systems, ontologies, semantic Web, and social and emotional computing (Ramos et al, 2009).

The fundamental idea behind the ITS is to (i) model the domain that is to be taught, (ii) deliver the training using automatically generated *teaching materials*, (iii) observe the training process undertaken by the student, (iv) model the student using the observations, (v) verify the *effectiveness* of the training by testing the student on the taught material – either continuously or periodically and (vi) create a streamlined *personalized learning curriculum* for each student. Hence, in developing an ITS, the goals revolve around using domain knowledge, understanding student behavior and teaching strategies for flexible individualized learning and tutoring. According to (Peter, 1999), the three core ITS technologies are (i) curriculum sequencing, (ii) intelligent analysis of the student's solutions and (iii) interactive problem solving support. On initial examination, it would seem that ITS would be a directly related and easily extensible system for our framework since both the systems are intelligent, model and adapt to the users and have a component that interacts with the users. But, on closer examination we notice that there are some core fundamental differences (to the point of making it a parallel system rather than a usable one) that exist between them. Table 1 examines these core fundamental features and their meanings in the context of ITS and the survey framework.

Feature	Intelligent Tutoring Systems	Intelligent Integrated Framework
Content	Requires extensive and complete domain knowledge to be generated, can be displayed in any ordering that conforms to	Questions are pre-defined by professionals from another domain (survey designers) and is subject to many rules and regulations in itself that the system cannot override, this

	some pre-defined rules and can be personalized for individual users	includes the ordering in the questionnaire models and rules regarding influence-able type questions in ATUS type systems etc.
Users	Students or learners. Users understand the objective of the system (to teach) and their own objective (to learn). Motivation and obligation exists highly.	Respondents. Users do not know/need to know the objective of the system fully and is limited to '*taking their opinion*', while their own objective is weakly defined to '*complete the survey*'. Motivation and obligation is minimal if it exists. Some might have more motivation such as "obligations to fill out the U.S. Census survey".
User Interactions	Bi-directional interaction. Users learn the domain content from the system while system observes and learns the student's characteristics. The system's objective is focused on the *insemination* of knowledge into the user.	Primarily one-directional or weakly bi-directional. The system may observe and learn the user characteristics while the system in the view of the user is only the means to complete the survey. The system's objective is focused on *extraction* of knowledge (or data or information) from the user.
Feedback	Exists and is intended to be uncontrolled. User's do not directly influence the system and the system has a certain degree of freedom on how the user's characteristics affect the learning and content delivery (fully or partially)	Minimally exists. Given the restrictions on the system, feedback cannot affect the actual content and must conform to visual cue rules and such for the system's GUI itself

Table 1 Comparison table for the different potential features in intelligent tutoring systems and the intelligent integrated framework

Thus, from the point of the integrated survey framework, the most adoptable feature of ITS research is the modelling of the user's performance by observing their behavior. This has been examined in a major way by (Cetintas, 2010). Here the author experiments with using simple observations of the student's interaction with the system, such as mouse movements (De Vincente & Pain, 2002) and time and performance features (Cetintas et al) to detect off-task behavior of the student. It must be noted here however, that there is significant other research in employing more sophisticated and dedicated equipment such as microphones, gaze trackers etc., but these do not comply with the framework's requirement and would be a hindrance in moving towards self-administration, where the system does not have any control over the client machine.

2.6.3 Recommender Systems

Recommender systems (RS) are applications of collaborative filtering research coupled with "extensive work in cognitive science (Rich, 1979), approximation theory (Powell, 1981), information retrieval (Salton, 1989), forecasting theories (Armstrong, 2001), management science (Murthi & Sarkar, 2003) and consumer choice modelling in marketing (Lilien et al, 1992), that help users deal with information overload and provide personalized recommendation content and services to them" (Adomavicius, 2005). An RS works with two primary entities – users and content. However, unlike the previously examined CAT systems, RS does not have an arbitrary scale for mapping defined. Instead, it uses different collaborative filtering logics to model both the users and the content simultaneously. The core objective of a recommendation system is that when the system is presented with a user u_1 who has interests I_1 ($i_1^1, i_1^2, \ldots$..) then the system must be able to predict what items from a set S would the user also '*like*'; the system then presents the selected items to the user and must verify if the presented items were '*liked*' by the user as the system had predicted. Thus RS attempts to model the users, use a recommendation process to determine the content that would best fit the user's model, present the user with the items, examine if the user's actual model conforms to the predicted model and apply corrective measures to the recommendation process itself in case of success or failure. Its many improvement features include better methods for representing user behavior and the information about the items to be recommended, advanced recommendation modelling methods, incorporation of various contextual information into the recommendation process, utilization of multi-criterion ratings, development of less intrusive and flexible recommendation systems that rely on the measures that are more effective at determining the performance of the recommendation system itself (Adomavicius, 2005).

An RS defines a utility function *u*, and works to predict *u* for a space defined by *C x S*, where *S* is the set of all the users of the system and *C* is the set of all the content in the system. The system may be provided with some utilities for some items in the *C x S* space. The predictions or extrapolations are done by specifying heuristics that define the utility function and then empirically validating its performance and estimating the utility function that optimizes certain performance criterion like RMS Error. This may be done using machine learning, approximation theory and other heuristics. An RS may work towards predicting absolute values for the utility value (known as ratings) or a preference based filtering prediction that is are relative preferences of many users. Most recommendation systems are classified based on the recommendation process as below (Balabanovic & Shoham, 1997):

i. Content-based

ii. Collaborative

iii. Hybrid

Content-based RS (CBRS) use the content as the similarity measure and have utility measured (for the user and the content) with respect to the content itself. Users are not modelled but the content is modelled using keywords that it contains and a textual search for recommendation items is done. This variation was heavily influenced by the information retrieval community (Yates & Neto, 1999; Salton, 1989) and as such takes a lot of contribution from them such as adaptive filtering, threshold setting etc. CBRS faces issues such as the limitation of content type to text (Sharhanad & Maes, 1995), overspecialization (where the recommended content may be the same topic/core worded differently; like a news report by multiple publications/sources), and new user problem (new users would have not rated anything yet and will have no utilities for any item in the content space).

Collaborative RS (CRS) essentially allow the users to model themselves by stereotyping them into groups. Say there is a group of users who rate content c_1 high and another new user c_1 with a high value. A CRS would now attempt to recommend other items rated high by the group to this new user. There are various approaches to implement this method such as the Grundy system, the Tapestry system, Memory based heuristics, Model based approach etc. CRS has the major advantage over CBRS that it can deal with any kind of content, since the content itself is not modelled. However, it still faces the new user problem now compounded by new item problem and sparsity (users need to rate sufficient number of items before being assigned to a group).

The third implementation is the Hybrid RS (HRS), which implements CBRS and CRS separately and then combines the predictions to create new recommendations. HRS may implement CRBS features into a CRS base system, or implement CRS into a CBRS base system or attempt to create a combined unified model in the Unified Probabilistic model. HRS use Bayesian Mixed Effects Regression Models (Markov Chain Monte Carlo) or Case-Based Reasoning for augmentation.

Recommendation systems are not directly similar to the integrated framework for automating surveys. However, the concept of user modelling and content recommendation is the backbone to the integrated framework for adaptation. Unlike CAT systems, RS allows for multiple dimensions (as multi-criterion) and grouping of users. The integrated framework equivalent of the C x S domain can be the space of respondents and their characteristics. It must be noted here that both surveys and RS face sparsity issues, but the integrated framework would have much sparser 'ratings' data. This would be induced since there are going to be many more states and gradations in the respondent's state and since a majority of the users would conform to a standard path, many of the states would be empty or have very few users in it. Also, in the case of RS, the items are within well-defined categories (such as genre) whereas in the integrated framework the user characteristics are more open to interpretation. RS also does not face the issue of simultaneously effected categories. In RS, items in Category A do not affect items in Category B, which is not the case in the integrated framework where user motivation value has an effect on the user interest value. For example, given that the framework might need to use percentage numbers for denoting the level of some characteristics like motivation and effort, the probability that there might have been another user with the exact same value for all the related (multiple) characteristics at the exact question through the same exact path could be slim. This situation would be compounded in a time diary survey where the user has the freedom to choose how to fill the activity responses up and the system considers the order of the survey as a matching attribute for recommendation.

2.6.3.1 An analysis of the opposing principles in Recommendation systems and Survey systems

The described research of the use of recommendation systems for survey systems would not be complete until specific attention is laid on the primary and ironically opposing principles in the two systems: Survey systems must strictly adhere to principles that define how bias is to be avoided and any form of influencing respondent decisions must be minimal (visual stimuli, ease of access etc.) while Recommendation systems are regarded as persuasive agents that recommend as well, according to (Gretzel and Fesenmaier, 2006). This persuasive potential in recommender systems has been increasingly observed in various works such as Häubl and Murray, 2003, Murray and Häubl, 2005, Bechwati and Lan, 2003, Bilgic and Raymond, 2005, Kramer 2003, Kruger et al, 2004, Mandel and Eric, 2002, Morwitz et al, 1993, Nass and Youngme, 2000. An important factor that has received comparatively little attention is the impact of the preference-elicitation process – the procedure used to capture users' likes and dislikes. According to the authors, this initial phase of the recommendation process creates expectations about the quality of the recommendations the system will provide, the structure of the preference-elicitation process and the cues the user derives from it can have a significant impact on the user's perceptions and evaluations of the recommendations.

Understandably, the paper describes the strong persuasive elements of recommendation systems – each of which corresponds to a problem with respect to survey systems where any and every form of persuasion must be eliminated. These elements are elucidated whenever encountered unless they are directly understandable.

According to Simonson, 2005 consumer preferences have been found to be susceptible to seemingly irrelevant factors like the set of alternatives included and the way questions about user likes and dislikes are asked. This implies that the recommender system plays an important part in the choice the user chooses using the system. The authors identify three important cues in the preference-elicitation process that are factors that influence users' perceptions of how well the recommendation matches their preferences as (1) relevance, (2) transparency, and (3) effort. The paper describes a metric named perceived fit which is defined as the user's belief that a recommendation represents an alternative that can satisfy his or her personal needs and wants. The paper goes on to describe and experiment with the factors identified as key factors. The paper finds out with statistical backing that the three factors are the significant ones but discover more factors that also play a part in the influence such as trust and cognition. The resulting graph is shown in figure 4.

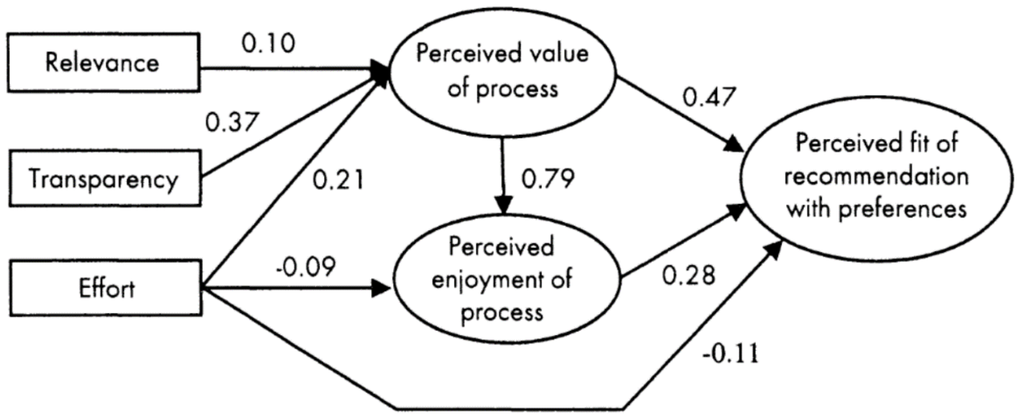

Figure 4 Significant Influences of embedded cues and perceptions of the preference elicitation process on perceived fir of the recommendations

2.6.4 Decision Support Systems

Decision Support Systems (DSS) are applications that help *evaluate potential decisions* by taking in all the information regarding resources that play in part in the selected *decision*. Adams R. (1990) claims that DSS can be seen as an extension of the idea of management information systems by providing a broader range of information in a more flexible and interactive way (Dawood et al, 2009). As such, these systems help in accumulating the information regarding factors (resources, facts, rules etc.,) into a set of decisions that can be used by human users (such as managers, officials etc.,) to examine their choices closely. Thus, DSS helps remove non-viable, restrictive, and time-consuming (if time is provided as a factor) decision options which is useful when there are too many available options to choose from. It must be noted here that any system/application that can consolidate data and filter them falls within the wide definition of a DSS. This means, for example, the Microsoft® Office Excel® application is a DSS when conditional filters are applied to eliminate mathematically and logically non-viable values (Power, 2000). Given the wide range of applications that fall within the definition of a DSS, some components exist common among them that infallibly form part of their *major components* such as (Dawood et al, 2009);

1. The end user – a decision maker(s)
2. A database/dataset source containing information of resources pertaining to the topic under the decision making process.
3. Models and procedures to simulate the effects of decision making
4. Module to manage the models, databases and the interaction between users and the system (GUI)

A variety of applications exists that use different methods to generate, select and simulate decisions such as simple filtering, sensitivity analysis (Pannell, 1997), Decision Tree Analysis (Apolloni, 1998), Cause-Consequence Analysis (de Meaux, 2008), Risk mode effects analysis and delphi methods (Hamilton, 1996 and Efstathiou, 2007), Analytic Hierarchy Process (Bareiss, 2004 and Andreica, 2009), Monte Carlo method (Damodaran, 2009 and Dey et al., 2002), Comprehensive analysis methods and Bayesian networks (Xiaocong et al, 2010). The applications themselves are used to analyze, simulate and generate plausibly efficient strategies, plans, layouts, risk analysis etc.

DSS, like RS, does not have a direct correlation to surveys in general. This statement of course precludes the scenario where the conductors of the survey use the data from the survey to evaluate and plan decisions. Since our integrated framework deals with the conduction of the survey itself, our statement stands valid. While surveys are a means to extract the opinion of the respondent, DSS is related to analysis of large quantities of data.

2.6.5 Information Retrieval

Information filtering or retrieval (IRS) systems are information systems designed for unstructured or semi-structured data; this is quite typical database systems which work on highly structured data such as employee records (Belkin, 1992). The idea of the structure used here is the way of formatting records – are they strictly defined (an employee record must have a name, age, identity number etc.) versus an email record (semi-structured data) which, while having well-defined header fields, also possess an unstructured text body. More often than not, information retrieval systems refer to textual data. Multimedia content such as images, voice and video are also often included under unstructured/semi-structured data for IRS. The process typically involves filtering incoming data, selecting relevant data (or elimination non-confirming data), Categorization and/or Selective Dissemination of Information (SDI) (Packer, 1979). Information filtering and retrieval are seemingly similar in their primary conceptualization and differ in that IRS is considered to have the function of leading the user to those documents that will best enable them to satisfy their need for information. The general model of an information retrieval system is given in the figure 5.

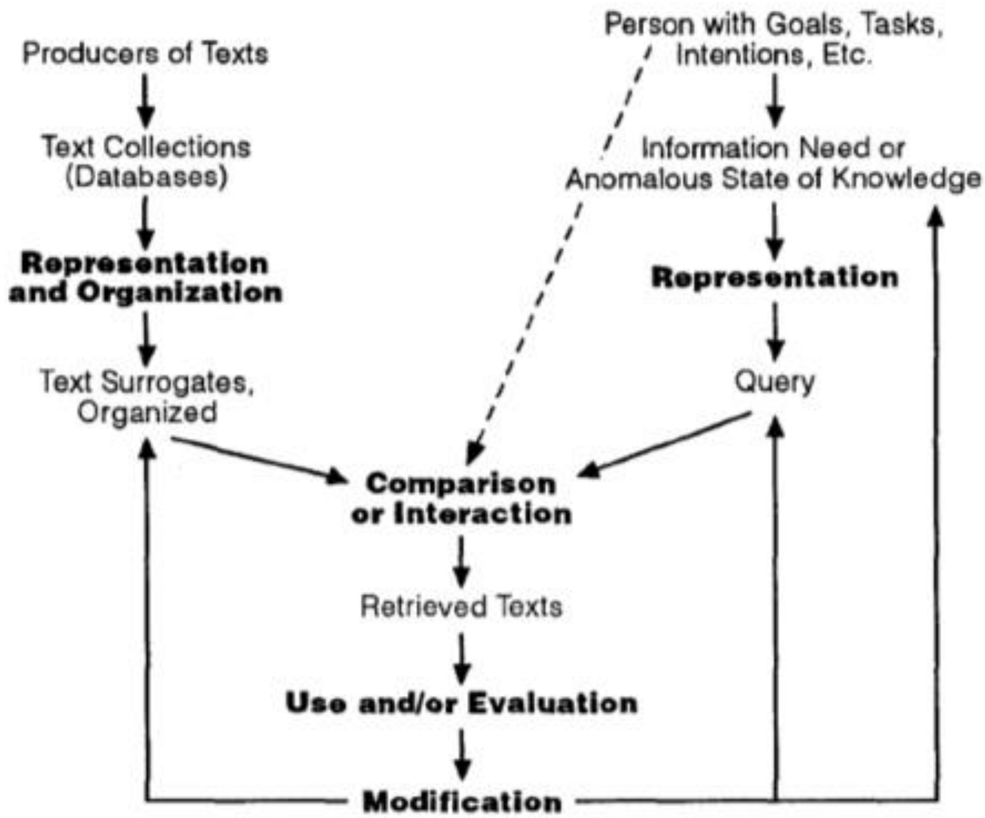

Figure 5 A general model of information retrieval (Belkin, 1992)

Based on the current literature review, IRS systems can be envisioned as descendants of text classification systems and in turn form a part of the backbone leading up to the different systems that has been described in the technology sections above such as CAT and RS. The three major comparison processes used in IRS are Boolean, Vector space and probabilistic retrieval models. While Boolean retrieval is based on an exact match principle, vector space and probabilistic models are based off of the concept of best match. Temporal constraints and its applications in IRS is an area of particular concern and attempts to understand when a text is likely to be timely for a particular user and when not.

As IRS is intended to work with either highly unstructured or semi-structured data and a direct correlation to survey systems is intangible, the processes used in IRS which has formed the building blocks for Recommendation systems and other processing technologies is worth our due attention. IRS usually employs classification, categorization of both users and data to intelligently assign resources and results to relevant users. Categorization, for example uses user profiles and models and assigns relevancy between user profiles and content. The parallel that can thus be drawn over survey systems is the assignment of prompts and probes to relevant users. This is one of core ideas that is expressed and attempted in this work.

2.7 Intelligent Learning and Reasoning Methods

While sections 2.6.1 through 2.6.5 examined the various technologies that use different learning and reasoning techniques for intelligent system design, there is also a plethora of methods and algorithms to infuse intelligence into a system. We consider a few of these methods in this section.

2.7.1 Reinforcement Learning

Reinforcement Learning (***RL***) is the method of learning behaviors using a trial-and-error interaction with a dynamic and uncertain environment (Kaelbling, 1996). It involves either searching the space of available behaviors to find one that performs well in the environment or using statistical techniques and dynamic programming methods to estimate the utility of taking actions in the world. It is widely studied in various disciplines such as control theory, game theory, Operations Research, information theory, simulation-based optimization, statistics and genetic algorithms. As such, there are many practical applications that use reinforcement learning as part of their intelligent behavior learning algorithms. The two main concepts that play a primary role in RL are **exploration** and **exploitation**. Exploration involves attempting to determine the possible effects of all available actions while exploitation involves targeting a known good state repeatedly by taking actions that are most guaranteed to lead again to good or better states. Assuming that there are some states **S** and some actions **A** that the agent can take, each transition **T** may be defined as a state change from $\mathbf{S_1}$ **to** $\mathbf{S_2}$ through some action **a** that the agent takes. A reward function **R** is associated with this transition that determines the effect of this transition on some objective that the agent is trying to attain. This objective is usually defined by some utility function **U** that the agent tries to maximize. In essence the concept of RL revolves around trying to find the 'right behavior' for an agent to best deal with an environment that it cannot completely control.

2.7.2 Case Based Reasoning

Case Based Reasoning (***CBR***) is a method of solving problems by retrieving relevant cases from specific previously stored case episodes and adapting them to fit new situations (Aamodt, 1994; Kolodner, 2014). It is modelled based on the natural method of anecdotal learning and much of the original inspiration for CBR came from the role of remindings in human reasoning (Schank, 1986). In its core essence CBS understands two facets of common human reasoning: (i) the domain of problems are regular that is, similar problems have similar solutions and (ii) problems encountered in an environment are usually recurring and not always unique (Leake, 1996). Given the complexity and richness of the nature of human behavior and reasoning, CBR draws motivation from it and the pragmatic desire to develop artificial intelligence. CBR works by having an initial set of cases – the prior knowledge or case base. The cases in this base set are indexed and described intricately to allow comparisons. When a new problem is posed, the system attempts to search through the cases in the case base and 'zero in' on the case that best fits the problem description of the new case. Using this case as the anchor, CBR now attempts adapt or use a trivially modelled solution from the existing solution to create a solution for the new problem. CBR can be extended to support learning and more complex adaptations. One of biggest challenges associated with using CBR in any domain is the design of the cases and the creation of the case base.

2.7.3 Cluster Based Modeling

Cluster-based User Modelling is a method of tackling the issues of sparsity and broadening the 'scope of search' in systems that model users and apply recommendations (O'connor et al, 2001). Traditional recommendation systems operate on individual user models to extract recommendations for new user models. This method suffers from the issue of nothing being able to 'cold start' and requiring an extensive dataset of initial mappings before being able to generate the recommendations. This is because initially, the system does not have sufficient user models properly defined to begin recognizing patterns for recommendation. Clustered user models are one of the many ways of handling this issue by grouping the user models into groups based on certain criterion. These grouping criterion may be predefined or adaptive. This effectively brings in a level of abstraction over the individual user models and reduces the sparsity significantly

(Ungar & Foster, 1998). In this method, the users and the items under consideration are from some classes or groups. Some of these classes may be predefined to accelerate the process of grouping. The system then effectively works on these groups, identifying relationships between the user groups and the item groups using methods such as repeated clustering or Gibbs sampling. Here the cold start situation is slightly mitigated since the space of search is now bigger (a group of items – aggregating the characteristics of all the items under that group) and would require lesser fully defined models. The issue with sparsity is also inherently addressed since now the same set of individual models have been reduced to a bigger sets of related models. Further refining process within the system could define more precise groups and increase the accuracy of recommendation at later stages when more and more user models become better defined. Here it must be noted that the integrated framework faces a similar situation when initially there may not be well defined user models and that the grouping criterion may be unknown. By employing the abstraction provided by clustered user models and the adaptive group criterion generation (using statistical methods), the integrated framework also addresses its issues in this regard.

2.7.4 Utility Theory

Utility Theory is a method of working with decisions by understanding and working with the concept of some 'utility'. The foundation of utility theory rests in domains where decision making is the expected outcome or objective (Fishburn, 1970). The fundamental theorem of utility may be considered to "do with axioms for preferences which guarantee, in a formal mathematical sense, the ability to assign a number (utility) to each alternative so that, for any two alternatives, one is preferred to the other if and only if the utility of the first is greater than the utility of the second" (Fishburn, 1970). This concept is built on the aspect of risk aversion and expectation of rewards. Simply put, when a utility is applied to a set of decisions, the best decision whether in terms of least risk or maximum reward must be the decision with the highest utility. Undoubtedly, this theory has been taken up for significant research in fields such as economics, business management, social behavior, psychologists, intelligent agent design in Computer Science etc. An extension of the utility theory is the 'Expected Utility Theory or Hypothesis' which deals with the hypothesis of an entity's (person, agent, group) preferences with regard to a set of choices it has with uncertain outcomes. It is generally agreed that the expected values can be computed by multiplying each possible gain by the number of ways in which it can occur, and then dividing the sum of these products by the total number of possible cases where, in this theory, the consideration of cases which are all of the same probability is insisted upon (Bernoulli, 1954). This mathematical function allows for defining a relation between expected value and probability, thus accounting for risk aversion behaviors. One of the most important work in the field of expected utility theory is the Von-Neumann-Morgenstern utility theorem which defines the criterion for assigning a utility function to preferences (Neumann et al, 1947). It defines four axioms (completeness, transitivity, continuity and independence) that if exists within a decision making setup, then a utility function can be applied to the decisions.

Considering the importance of utility theory in decision making and the integrated framework's core need to make decisions (to improve responses or reduce break-off), it makes utility theory and the concept of ‘utility’ a necessary and vital method to consider. For example, in CBR, the cases could be assigned utilities (which are calculated after observing the result in those cases). This utility can be used by CBR to select the appropriate similar case based on the current environment. Utility could be assigned to various cues, prompts that could be maintained using RL by observing and learning the effects of those cues or prompts. Probabilities can also be incorporated as weights for these utilities, justifying further the use of expected utility theory.

2.7.5 Relevance Feedback

Relevance feedback is an implicit feedback technique that is a very attractive candidate to improve data retrieval and recommendation performances (Hill et al, 1992; Kamba et al, 1997; Morita and Yoichi, 1994; Seo and Zhang, 2000). Implicit feedback techniques gather data indirectly from the user by monitoring behaviors of the user during and after searching. If the information about search results’ relevance to users’ queries can be gathered passively rather than actively, then users can experience the benefits of relevance feedback without having to expend any additional effort – which is an extremely desirable feature in the survey system since the respondent and the interviewer’s cognitive attention is better suited to be focused on the survey response. A wide variety of relevance feedback techniques exist for analyzing web based and document search results, though it’s use in survey systems and particularly time diary surveys would more likely be directed towards relevance feedback in application user feedback.

2.8 Summary

This chapter provides background information and the literature review that went into shaping this work and defining our integrated framework from both the survey point of view and the Computer Science point of view. By understanding the requirements of the survey domain (particularly time use surveys) before perusing the technologies available to creating an intelligent integrated multi-mode time diary survey framework, we are able to better understand the complexities involved and why research in this direction has been slower when compared to other domains. This sets up the integrated framework with an ambitious final objective, and this work as the first few steps in that direction. By addressing the primary problem that surveys need – a human component, our integrated framework attempts to bridge the gap towards creating fully functional intelligent survey instruments that could completely replicate interviewer behavior.

Chapter 3: Methodology

3.1 Overview

The framework designed and developed in this thesis aims to provide a single intelligent integrated time diary survey framework that can be used with minimal modifications and effort by two different types of users; respondents and interviewers. Current work in this direction has been limited and the concept of integrating has not been tackled leading to the development of different survey tools for the different types of users. However, this comes with issues further down the line in the survey domain since the survey data obtained from multiple sources need to be homogenized for comparative research. While handling the issue of homogenization isn't the intention of this thesis, we take a step in this direction by creating an integrated framework that can be modified and adapted to suit the needs of the user. The demands on a survey instrument are numerous and mostly driven by the need to generate good data. Since the users interact with the survey instrument, the instrument plays a part in inducing expected or unexpected behaviors in the user which in turn affects the quality of the data obtained. Our thesis thus lays the ground work and expounds on the characteristics, the problems faced and the solutions to creating an intelligent integrated multi-mode time diary survey system.

The objective of an intelligent integrated time diary survey framework is to enable elicitation of the required information from the user in a manner that keeps the user engaged while providing assistance to the user to enable them to interact with the instrument with ease. The added complexity of a multi-mode behavior, wherein the user can either be the respondent itself in a self-administered mode or the interviewer in an interviewer-assisted mode, brings about different priorities depending on the user. While the integration of the two modes would cursorily seem to be two different problems, we attempt to unify them as simply users with varying intentions, motivation and knowledge. Thus our framework would take a highly abstract view of the problem of building an integrated multi-mode time diary survey instrument enabling us to leverage the characteristics of a user type to handle the delineating characteristics of the alternate user type.

Conventional work along this domain as described in Chapter 2, looks at the two different types of users as distinctly separate where a self-administered instrument would essentially be significantly different with the instrument attempting to simply replicate the actions of an interviewer through case based reasoning or reinforcement learning using a set of defined cases or rules. Our framework diverges from this approach while still maintaining many aspects close to or similar to the existing methodologies. By integrating the two user modes and using the Internet as the platform, we increase the accessibility of the instrument. In the modern scenario, where the Internet and the use of web applications has reached new heights, a survey instrument that employs the web can target users that would otherwise seem unreachable.

The separation of the two users would have brought about the design, development, and maintenance for two different instruments in a conventional scenario. By integrating them, we attempt to provide a generalized solution since we presume that there would be significant parts of the two individual instruments that would be similar in purpose, function or code. Our framework thus chooses to integrate the two user modes to handle this from the onset itself. Furthermore, our framework views the task of integrating the multiple modes of administration as its primary objective and thus our work probes into what makes the two modes different and how this difference can be resolved in a manner that leverages information and the characteristics of one mode and uses it to handle the problems encountered by the other.

3.2 The underlying principles

Understanding what is expected from a survey, the advantages and disadvantages of F2F and web-based surveys as described in Chapter 2, we describe how the integrated framework works in this chapter. In the process of describing the framework, we use the application to the survey domain to help describe the ideas and discuss the issues addressed in the framework.

First, the following lists the broader set of rules that shapes this framework.

- **Rule 1: User Assistance**: The framework must actively work to assist the user (respondents/interviewers) in recording their true responses.
- **Rule 2: Minimal Modifications between Modes (MMbM)**: Must work with zero to minimal modifications between interviewer-assisted and self-administered mode.
- **Rule 3: User Type Agnostic in Design**: Must be capable of interacting directly with both types of users: the respondent (self-administered mode) or with the

interviewer. This does not imply that the instrument cannot take into account the type of the user, but simply that the instrument must use its interaction mechanisms to cater to them differently without requiring specific designs for the two types of users.

- **Rule 4: User understanding:** Must observe the respondent's and interviewer's behavior and learn to model them using paradata.
- **Rule 5: Knowledge Engineering Phase:** May require a separate knowledge engineering phase with a dedicated/motivated human entity, but ideally should be able to understand data on-the-fly with a short starting phase.
- **Rule 6: Adaptation:** Must use the modeled user behavior to facilitate adaptive designs, for example, predict, detect and mitigate possible (if not all) survey-related issues such as break-off, socially desirable responding, lack of motivation etc.
- **Rule 7: Non-influencing entity:** Must not influence the respondent's opinion or suggest ideas consciously or sub-consciously to the respondent. This means that the instrument must not lead or bias the respondent to pick a specific option (recommendation) by making it easier (lesser effort) compared to the respondent's true response.

The framework is intended to lay the foundation to building fully automated intelligent self-administered survey delivery systems. However, on examination, one can realize how this framework is effectively attempting to address domains that require similar automation of human-to-human interactions for knowledge extraction.

3.3 Problem Description

The first step in designing and building our framework is understanding the domain problem of the multi-mode time diary survey. This involves describing the environment of the two modes, the related modeling problem and finally the interaction of the environment and the users. The necessity of modeling the environment, the users and their interactions comes from the fact that framework acts as a conduit between the user with the information and the elicitation and recording of this information. Thus the framework must understand the characteristics of each user type and how they are similar and different. With the user modelled, the framework must then understand the environment that the user exists in and how the user interacts with the environment. The framework can then interact with the user in such as a manner that it assists the user in eliciting the required information reducing their cognitive burden that comes with time diary surveys.

3.3.1 Data (Modeling) Problem Description

In this section we describe the different aspects of the time diary domain problem. We discuss the challenges of integrating the two administration modes and the inherent characteristics of each mode and their unifying aspects. We then build our framework with an abstract standpoint that can then be reduced to a finer and more implementation oriented standpoint.

3.3.1.1 **Survey Modes Modeling**

Time diary surveys are intended to elicit information about the respondent's activities for a given time period (4 am the previous day to 4 am on the day of the interview in ATUS). The information includes the activity performed with their starting time and ending time and contextual information such as who they were with and where they performed the activities. When more than one activity is reported by the respondent for the same time period, one of the activities is regarded as the primary (or main) activity with the other activities being secondary activities. The selection of the primary activity is usually provided by the respondent itself based on their personal discretion. In the self-administered mode (SAM), the respondent directly interacts with the instrument and thus have to recall their activities and record them using the instrument on their own. In the interviewer-assisted mode (IAM), a trained interviewer acts as an intermediary between the respondent and the instrument and guides the respondent through the recollection process while recording the activities in the instrument. These interactions are illustrated in Figure 6.

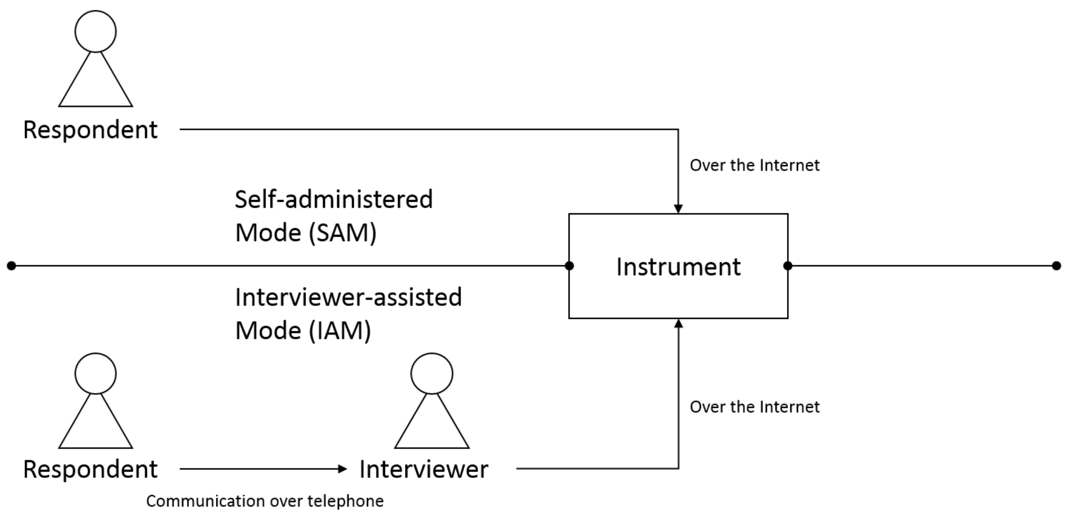

Figure 6 Illustration of the user interactions for the two modes

3.3.1.1.1 Interviewer-assisted mode

When the user using the instrument is an interviewer, the instrument is said to be operating under the interviewer-assisted mode. This setup is similar to that of Computer Assisted Telephone Interviews (CATI). When operating in this mode, the respondent characteristics are unavailable to the instrument directly. The interviewers logging data (also known as paradata) is available and so is the response data as recorded by the interviewer. It must be noted here that the only information regarding the respondent that is available in this mode is the response data and even so, the response data is not a direct indication of the respondent characteristics as it is the interviewer that records them and is thus influenced by the interviewer's characteristics also.

3.3.1.1.2 Self-administered mode

When the user using the instrument is a respondent, the instrument is said to be operating under the self-administered mode. This setup is similar to an application used by the respondent wherein the usage is fully controlled by the respondent themselves. In this mode, since the response data is directly recorded from the respondent, it is, together with the paradata, information directly relating to the respondent characteristics. However, under this mode, the respondent has complete control and discretion on using the instrument and conversely their participation is directly influenced by their understanding of using the instrument.

3.3.1.2 **Interview Modeling**

This section describes the characteristics of the survey (interview) itself. The term interview is appropriate in the interviewer-assisted mode while the term survey is appropriate in the self-administered mode. However, the two terms simply denote the process of conducting the survey and are used as such. As our integrated framework attempts to deal with the two modes as one mode with variable user characteristics, the problem description here is that of what the differences and similarities are. The similarities would constitute the user agnostic aspects while the differences constitute the user specific aspects.

3.3.1.2.1 User agnostic aspects

The user agnostic aspects of the survey are described by the data obtained throughout the survey process. The data represents the observables of the environment and is obtainable independent of the user and the administration mode. Each of the data can then be used to infer the possible characteristics of the user by the instrument which can then be used to change the behavior of the instrument to best fit the scenario.

3.3.1.2.1.1 Response data

The response data encompasses the information recorded by the instrument pertaining to the response provided by the user. This includes the actual user response to instrument's questions, the mapping of those responses to system-identified responses and finally the process for the execution of such mapping. In our implementation, user response to instrument's questions is also known as "activities" as user provides an accounting of their daily activities for the time diary. Meanwhile, the system-identified assets are collectively known as auxiliary data and constitutes the data that is used by the system to understand the responses provided by the respondent. This enables the system to thus *identify* the activities reported by the respondent allowing it to use the information. This information can be used to assist the user by learning and adapting to the knowledge contained within the system.

3.3.1.2.1.2 Paradata

The term paradata is attributed to Couper (1998) and is an overreaching term to contain the administrative data about the process by which the survey data was collected. In the view of the Computer Science domain, this is closely related to what is known as logging data. Examples of paradata include the length of the interview, the observations within the interview process such as how the data was entered and edited and the methods by which the data was modified. Together with the logging data, this also extends to how the user interacted with the system – mouse movement, keystrokes etc. Thus paradata is usable as being indicative of the user characteristics that determine their understanding of the instrument, the survey and how it influenced their interactions with the instrument.

3.3.1.2.2 User specific data

User specific data is the data that distinctly separates the respondent using the instrument from the interviewer using the instrument. This influences how the user agnostic data can be interpreted and is hence highly tacit. For example, the respondent's interaction with the instrument is influenced by their understanding of the purpose of the survey, their motivation to sit through, recollect and record the response data in a manner that makes most sense to them. On the other hand, the interviewer is a trained user with a firm grasp of what information to collect, how to collect it and how to record the information. Thus the user specific data would be the interpreted data based on the user agnostic data obtained. This is thus a cornerstone of the integration process wherein, both SAM and IAM can exist within the framework with the distinction being drawn only as internal data. This eliminates the need to handle SAM and IAM as two different modes since only those uniquely specific data that is inferred needs to be handled differently. For example, when the relevant context information is missing in SAM, the instrument has to probe the respondent for this missing information in an appropriate way so as to reduce the respondent's burden and increase the response content. In case of IAM, this missing information may be indicated to the interviewer (e.g., missing fields indicator) and thus the process of obtaining them is deferred to the interviewer.

3.3.1.3 **User Modeling**

User modeling generally involves fitting the characteristics of the user under a set of predefined attributes. These attributes can be the user’s skills and/or their declarative knowledge. The main goal of user modeling is to customize and adapt the system to the user’s specific needs, thus allowing the system to ‘say the right thing at the right time in the right way’ (Fisher, 2001). As user modeling typically involves assigning the user to certain values within a scale (which may be continuous or discrete), the entire range of possible values of the scale must encompass all possible values attainable by any user of the system pertaining to the defined purpose of the system. In case of time diary surveys however, this distinction would essentially separate respondents from interviewers quite distinctly and hence current literature and related works look at respondents and interviewers differently. In cases that attempts to handle both of them (the ATUS instrument by census), one of the user becomes the primary target (the interviewer in ATUS), with the other user (the respondent) having to adapt themselves to use an instrument that is not uniquely tailored for them. While this would not be a severe issue in a system where a user uses it for their own benefit; in time diary surveys, it becomes a source of frustration for the respondent since they stand to gain no direct benefit from it resulting in them simply quitting.

Our integrated multi-mode framework views the user as an abstract entity focusing on the source of the actual information, which is always the respondent. Thus in SAM, the respondent directly interacts with the system and hence the instrument has direct access to the respondent. In IAM, the respondent communicates the information to the interviewer and the interviewer interacts with the system and thus the instrument can refocus on assisting the interviewer in eliciting the information from the respondent who is the source of the actual information. This then allows us to model the user based on their motivation in revealing the required information and their expertise in recording this information in the instrument. These two factors (motivation and expertise) allows us to distinguish the two types of users where required while considering them as users.

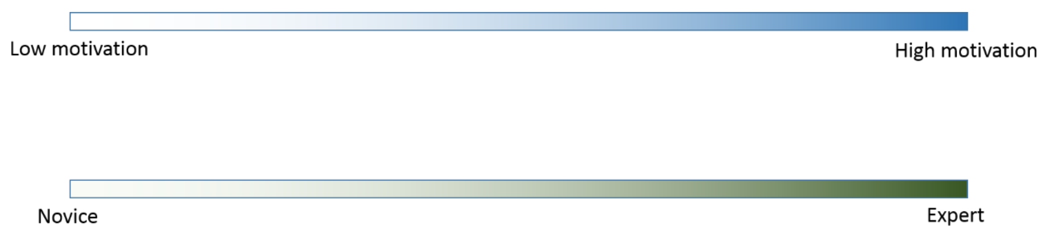

Figure 7 Motivation and Expertise Scales

Error! Reference source not found.7 shows the continuous scales related to m otivation and expertise. From this we extract the four end points and Table 2 details the characteristics that is to be expected from each of the four.

End Point	Characteristics
Low Motivation	Users with low motivation would tend to attempt to complete the interview as fast as possible without having to exert significant cognitive load. When this is not possible, low motivation users can be expected to quit or get frustrated.
High Motivation	Users with high motivation would attempt to complete the survey and can be expected to put in the effort required to understand and learn to use the instrument to fulfill the requirements.
Novice	Novice users are characterized by their lack of knowledge in using the instrument. Their actions during their encounter with the instrument would be chaotic and subject to high amounts of trial and error. They would try to click and observe the functionality of the instrument before delving into the survey aspects itself.
Expert	Expert users are familiar with using the instrument and can be assumed to be mostly precise in their usage. Knowing what information is required and how the information is to be entered into the instrument would enable them to focus more on the survey aspects rather than on dealing with learning how to use the instrument.

Table 2 Characteristics features of the four end points from the motivation and expertise scales

While motivation and expertise have been provided as the separating features, current research in identifying and measuring motivation and expertise is limited and non-existent in the field of time diary surveys. However, there have been indicative research findings that point to how cognitive loads, response completion and satisficing during surveys are affected by motivation, fatigue and expertise (Krosnich, 1991; Backor, Saar & Norman, 2007). Fatigue has been linked to reduced data quality, while motivation has been related to increasing response rates and reporting.

3.3.1.3.1 Interviewer Modeling

Interviewers conducting the interviews, as mentioned earlier, are focused on keeping a conversation with the respondent. Through this conversation they extract the information required from the respondent. Once they are privy to certain information, they enter the information in the instrument. As trained users, the interviewers can be assumed to be highly motivated users. With respect to their expertise in using the system, however, they can range from being novices to experts since their understanding and learning of how to use the instrument, changes as they use the instrument more and conduct more interviews. Their expertise in conducting interviews are however beyond the scope of this thesis, but indicative measurements may be obtained by comparing the aspects of the interview such as time and the quality of the data obtained. Their expertise in using the instrument however can be observed and studied closely by analyzing the paradata collected during the interview process and analyzing it besides the quality of the data obtained.

3.3.1.3.2 Respondent Modeling

Systems employing user modeling generally consists of identifiable user features that can distinguish and/or identify users distinctly or in groups. For survey systems however, this feature generates a new challenge since the data must be de-identified off of all user information. There may or may not be repeat users over time and these users may or may not be allowed to possess identity features (such as a unique username or number). Different survey systems employ varying degrees of stored user identification information such as case ids, respondent numbers etc. Our framework bases off on the assumption that there will not exist any directly identifiable respondent information available to it for use as minable data. This does not include user information stored for the sake of keeping track of the interviews scheduled/completed. This adjustment is necessary for the sole purpose of making the survey instrument accessible securely over the Internet with features such as resume later; however, none of the user information will be used by the framework for knowledge engineering or analysis thus allowing this information to be pseudo-generated keeping privacy issues at a minimum.

The aforementioned inability to identify respondents uniquely brings to the table the issue that the framework cannot assign information to particular types of users. This is however not an issue when considered from the point of view of motivation and expertise since depending on known features, the respondent may be assigned an arbitrary starting motivation and expertise which the system can then either adapt as they progress through the survey or keep constant. While it may seem intuitive to label respondents as unmotivated users, certain types of surveys and respondents are generally motivated; for example, ATUS panel respondents who have been involved in time diary surveys for longer periods of time can be assumed to be motivated considerably. Thus assigning and managing the motivation for the respondents would require future research and work.

Unlike interviewers however, respondents could face a significantly more challenging issue with learning to use the instrument; that is their expertise. This is further added on to when considering that respondents need to effectively perform both the recall and the record actions themselves leaving little space for learning to use the instrument effectively. Figuring out how much help the respondent would need with the instrument is thus essential and must be obtained as soon as they begin (or before) the survey. This can be accomplished by a simple questionnaire regarding their previous experience using the instrument and later followed up by using the paradata from the survey session.

3.3.2 Interaction (Modeling) Problem Description

The term “interactions” refers to both actions and information that is passed on between the instrument and the user. Thus it includes the information text presented to the user by the instrument on one end and the user’s response to some information presented on the other. By placing emphasis on these interactions, the framework can attempt to identify and adapt to changes that may be derived by observing the interactions.

3.3.2.1 **Interviewer –Respondent Interactions**

The interactions between the interviewer and respondent in IAM are the hardest to capture and measure and in our study ignored. Since the interviewer is the final human entity interacting directly with the instrument, any data available to the instrument would be painted by the interviewer's interactions with the instrument rather than the respondent. However, the interviewer-respondent interactions would partly be responsible for how the interviewer records the data; if the respondent is slower in recalling and responding, an effect of this should be a decrease in the speed that the interviewer records the activities. Thus these interactions can be used for identifying respondent's characteristics in IAM which can later be transferred across to SAM to deal with similar respondents.

3.3.2.2 **Interviewer – Instrument Interactions**

When the interviewer interacts with the instrument in IAM, they are essentially acting as a conduit between the respondent and the instrument. Their role in this interaction is enormous since the interviewer largely controls the interview process. They transfer the information provided by the respondent to the instrument while also eliciting the said information from the respondent through queries and probes in conversation. When interacting with the instrument, they enter the information provided by the respondent either verbatim or apply human reasoning to fit the responses to certain defined survey standards. They may use features supported by the instrument to aid them in entering the information faster and in reducing the errors entered. Thus this interaction can be viewed as an exchange of information between the interviewer and the instrument wherein both of them have the same goal of creating good quality survey data. Furthermore, interviewers are more likely to respond to the instrument and can be expected to take the trouble to understand any issues with the instrument.

3.3.2.3 Respondent – Instrument Interactions

Respondents would be directly interacting with the instrument when the instrument works in SAM. The interactions between the respondent and the instrument are likely to be more capricious since the respondent has full control of the survey process. Their interactions can vary between trial-and-error situations as they figure out how to use the instrument, to more refined usage scenarios where they are attuned to using the instrument. They may respond with hostility (break-off) or may welcome information presented by the instrument. Thus all interactions directly with the respondent must be controlled and balanced; not assisting at all would be just as bad as putting words into the respondent's mouth. This is further limited by the non-influencing entity rule (Rule 7) described in Section 5.2, wherein the timing of assistance plays a role.

From the respondent's point of view, the instrument should be easy to use, intuitive and reduce their cognitive load as much as possible. Keeping things interesting would be added plus. From the instrument's point of view, it is metaphorically replacing the interviewer and must perform the tasks that would have been otherwise performed by an interviewer. This includes explaining the survey process to the respondent, guiding them through the survey and assisting them in recalling and recording their responses. Thus the interactions between the respondent and the instrument need to be simple, succinct and timely.

3.3.3 Problem Summary

Thus, in essence, the problem can be described as creating a time diary survey instrument tool that can be used by both respondents directly, and by interviewers in a CATI setup. Challenges in creating a solution arise first from the limitations imposed on the instrument for use with time diary surveys. The design must be consistent and the framework must not introduce negative effects on the users. Secondly, while general survey instruments such as questionnaires have made progress in being web friendly, time diary surveys have not made significant leaps in the same direction. While research in time diary surveys is limited to address either the interviewer or the respondent from a survey methodology point of view, no significant efforts have been introduced to attempt to bring the two modes together. Integrating IAM and SAM into the same framework allows us to create one tangible product capable of delivering time diary surveys to interviewers and respondents with little difference in the time between the development of each. This would also enable generation of consistent data for both the modes, with the same implementation running consistently on the platform it was designed for. This would reduce the complexity involved in switching and adding new features and eliminate the need to perform these changes on two separate implementations. Furthermore, given that interviewers and respondents approach and use time diaries differently, the instrument catering to both must effectively be able to switch accordingly. Thus a single integrated multi-mode time diary survey framework sets up the way in building an instrument that can deliver time diary surveys over the Internet, help the respondent or the interviewer in completing their time diary while being easily deployable and modifiable. All these features would be able to make the task of conducting time diary surveys a more approachable task. Provisions must be also made for the easy implementation of solutions offered by the different methods and technologies discussed in Chapter 2. This would allow the framework to expand and integrate future implementations of intelligent components with minimal modifications.

3.4 Proposed Solution

3.4.1 Abstract Framework Description

The proposed framework is aimed to setup an understanding of the survey domain in the context of modeling instrument and user interactions. These interactions are in two forms: Between the instrument and the interviewer and between the respondent and the interviewer wherein the interviewer uses the instrument to record the respondent's responses. One aim of the framework is to replicate the interviewer-respondent interaction in a Computer Assisted Telephone Interview (CATI) setup where the interviewer assists the respondent in completing the time diary survey (TDS); the framework would provide assistance to the respondent directly taking on a role similar to the interviewer. The framework is also tasked with providing assistance to interviewer when used as the instrument in a CATI setup, where it assists the interviewer to focus more on the communication with the respondent rather than on the menial task of recording the responses. As seen earlier, when the respondent directly interacts with the instrument, the instrument is also tasked with assisting the respondent to focus more on providing the true response rather than on learning and figuring out how to use the instrument; a role performed by the interviewer when they assist the respondent. This distinctly creates two modes of operation: (1) the interviewer-assisted mode (IAM) where the core purpose or aim of the framework is to assist the interviewer in recording the respondent responses and reduce the interviewer's cognitive burden regarding the same, and (2) self-administered mode (SAM) where the core purpose or aim is to assist the respondent in using the instrument to record their responses and reduce the respondent's cognitive burden when using the instrument. These tasks involving different aspects of modeling the interview, the users (respondents and interviewers) and the ways the models can be leveraged to provide the required assistance. Given that there are two different types of users, current literature shows that the two users are always handled differently as in, there are instruments that cater to interviewers specifically (like the instrument used for CATI) and there are separate instruments used to deal with respondents (like web forms) (Couper, 2000). While it makes sense at the implementation and research level to tackle the two users differently, our framework's broader approach allows us to view this distinction in terms of different metrics such as the user's purpose and motivation, and system usage knowledge. This brings about the core understanding of the integration-based approach of our framework. By making the framework be user agnostic in design (Section 3.6, Rule 3), we effectively move the concept of the type of user from the instrument's perspective into the framework's perspective. Thus while a user uses the instrument, depending on whether they are a respondent or an interviewer, different mechanisms kick into place that use 'user data' (again this depends on the mechanism) to cater to their corresponding purpose, motivation and system usage knowledge. This gives us a two-fold advantage: (1) By bringing about the separation of the users at the framework level, our instrument is user agnostic in design i.e., the mechanisms switch rather than the entire instrument. (2) Usability of expert knowledge systems, where the experts can be expert level interviewers or expert level respondents allowing us to draw the required knowledge from two different types of experts. This brings about the full circle of our framework's integrated, multi-mode approach.

We begin by describing the core mechanisms that the framework requires. For this framework, we define two core sets of mechanisms that would enable the framework to deliver on the various rules (Section 3.6) laid previously. Each set of mechanisms describes the environment within which it exists, the problems encountered and subsequently the solutions that fit in the environment of the mechanisms. This brings about the fundamental picture of the framework as shown in Figure 8.

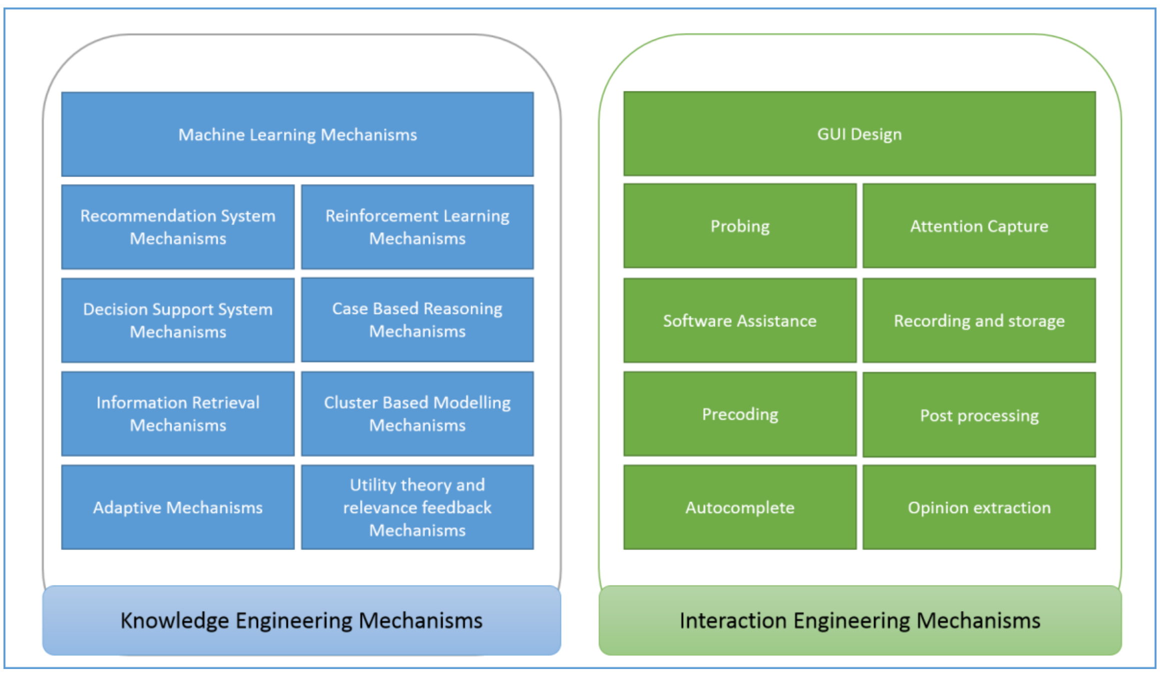

Figure 8 Abstract representation of the Integrated Framework using solution mechanisms

The idea of bringing everything in the framework under two broad sets of mechanisms allows us to view many of the problems mentioned in Chapter 2 in the context of each of the mechanisms. Each of the mechanisms contributes to handling one particular problem aspect of the domain – thus allowing multiple mechanisms to be coupled together to end up building a completed framework.

The two primary sets of mechanisms of the integrated framework are:

1. The Knowledge Engineering Mechanisms, and
2. Interaction Mechanisms

3.4.1.1 **Knowledge Engineering Mechanisms**

Knowledge Engineering (also known as Knowledge Modeling) (elicitation, analysis, construction, representation, implementation, validation, and maintenance) is what we call the process of knowledge elicitation, representation and management. The process of knowledge modeling can be broken down into two major tasks: initial knowledge modeling and knowledge maintenance (Aamodt, 1995). While the initial modeling phase involves knowledge elicitation, analysis, construction, representation and implementation, once the system has moved past the initial state, the process involves the validation, management and maintenance of the knowledge.

The Knowledge Engineering Mechanisms relate to the handling of various problems and issues that arise primarily when considering the data (knowledge) of the domain of time diary surveys. Various knowledge engineering methods currently in use in other domains have been mentioned and examined in Chapter 2 with their advantages and disadvantages for use within the framework. Knowledge Engineering Mechanisms for the framework consists of all the mechanisms by which the system gathers (including recording and parsing), processes (cleaning mechanisms) and maintains access to relevant data. As such, we see that the knowledge engineering mechanisms fall under two specific categories based on its running conditions. **Online** mechanisms are those that are active during the time of use of instrument (otherwise known as live system) and that can actively use the data being collected (before or after cleaning and processing) to improve the system in real time. Thus online mechanisms help in bringing about feedback and reinforcement mechanisms into the system. Understandably, not all data from the live interview may be accessible depending on the survey being considered or the data may be too enormous to be used or kept as resources in the live system and this brings about the need for **Offline** mechanisms that execute when the system is not in use. Offline mechanisms help in handling issues mentioned in Chapter 2 such as the cold start problem and scalability of data issues. This is analogous to offline and online learning in other applications of learning tasks, for example in reinforcement learning as in the work of Sylvain & Silver, 2007.

3.4.1.2 **Understanding Knowledge Engineering**

Knowledge engineering as defined in Section 3.4.1.1, deals with the data in the domain; in our case this falls under two types:

1. The interactions data between the interviewer and the instrument, and the interactions data between the interviewer and the respondent.
2. The response data that is recorded by the instrument.

Table 3 lists the environment for the data for knowledge engineering is based off of the interactions between the interviewer, instrument and the respondent. These interactions are characteristically different from one another as will be described in Section 3.4.1.2.1. Once an understanding of the environment is established, we examine the issues faced when designing knowledge engineering mechanisms in Section 3.4.1.2.2 followed by our solutions to these issues in Section 3.4.1.2.3.

Environment	A human- to- human interaction for the purpose of extracting knowledge from one willing participant by another. The presence of knowledge with one party does not make that party an expert participant, instead makes that party the only source of this required information, with no alternate source of validating the same.
Issues Expected	Unknown true response, interaction complexity, understanding the loss of one of the participant in the interaction, cold start, drawing relatable data from the interaction
Solutions Available	Use existing data to create adjustable baselines, view the data differently and within the context of one particular problem, apply mechanisms depending on available resources (online if priority is adaptation, offline if priority is access to information)

Table 3 Understanding what defines the Knowledge Engineering Mechanisms' purposes

3.4.1.2.1 Knowledge Engineering – The Environment

In a TDS, the respondent is asked to recollect the activities they did during a period of time together with contextual information such as who they were with and where they did the activity. While traditionally this was self-reported by the respondent using a pencil-and-paper method, we are concerned with the more recent interviewer-assisted method. In this method as in a CATI system, the interviewer would call up the respondent on the telephone and communicate with the respondent asking them to recall their activities and record them using a software system (instrument). Thus the following interactions exist in this setup:

1. Interviewer-Respondent interaction

This is a type of human-human interaction. The interviewer explains what is required from the respondent and may provide an example for the respondent to understand. Once the respondent understands the purpose of the interview, they begin their recollection process and tell the interviewer the activities they performed that they remembered. This may or may not be in a chronological order. Depending on the interviewer's discretion and the instrument's limitations, the interviewer may guide the respondent to go in a chronological manner using techniques such as backtracking (where they repeat the previously reported activities and try to help the respondent remember what they did next) and visualization (where they help the respondent visualize their day and help them recollect). The interviewer would also try to help the respondent correctly recall the required contextual information for the activities they perform. The respondent's responses to the interviewer can be highly varied and unstructured. They may also be cooperative or uncooperative, good at recalling or bad at it and hence their responses would be affected accordingly.

In this interaction, the interviewer forms a picture of how best to help the respondent. They may patiently explain to the respondent what they need and ask them follow-up questions to guide the respondent. The main type of data that is extractable from this interaction is thus about the respondent:

a. Are they cooperative or uncooperative?
b. Are they able to recall or not?
c. Do they prefer reporting what they remember first or in chronological sequence?

d. Will they complete the interview or break-off?

When modeling this interaction, the framework thus must attempt to identify the issues related to the above data regarding the respondent.

2. Interviewer-Instrument interaction

This is a type of human-computer interaction. The interviewer uses the instrument to record the information provided by the respondent. Currently, as explored by Chapter 2, the instrument tends to be a ‘dumb’ software where it simply records and provides basic validations and at times rule-based prompts to the interviewer. Since the interviewers are trained to use the software, this interaction essentially represents:

1. The understanding of the instrument by the interviewer – are they an expert (they are adept at using the instrument) or a novice (they are new to using the instrument)?
2. Does the instrument provide the interviewer with all the data they need to help the respondent?

Understandably, the interviewer’s interaction with the instrument is also influenced indirectly by the respondent. The interviewer may not be able to record information because the respondent may not be providing the required information.

To summarize the environment for knowledge engineering, we have two parties (interviewer-instrument, respondent-interviewer) that are engaged actively (as for interviewer-respondent) or passively (as for interviewer-instrument) in the process of elicitation and recording of response data. The knowledge of the system thus is a representation of the possibility of successful completion of the interactions and of the validity of the data obtained and recorded through these interactions. The key here is the realization that whether the user is a respondent or an interviewer, they essentially must be able to use the instrument (when the respondent also uses the instrument directly in SAM) effectively to deliver their intention which is to record the data. The validity of the data may not be verifiable in any case since there are no alternate data points of the respondent to verify it (and verification may not even be necessary since these are subjective information pertaining to an individual – respondent). The instrument's purpose is thus identical regardless of the user; it must simply suit itself to the user's disposition to perform the interaction.

3.4.1.2.2 Knowledge Engineering –The issues faced

As seen in Section 3.6.2.2, the environment for knowledge engineering is related to the interactions between the three parties involved – the instrument, the interviewer and the respondent. The fundamental aspects of knowledge that needs to be extracted are the user's motivation, purpose and their knowledge of how to use the system (the system usage). Unfortunately, there is currently almost no literature that defines these terms in terms of time diary surveys. However, these terms are not completely new when looking at them from the point of view of software systems. Motivation is primarily seen as the drive to perform a task or objective, for example, the motivation of users in knowledge management systems is to contribute to the system by creating, sharing and using the knowledge within it (Malhotra et al, 2003). When motivation is high, it is expected that the users are 'motivated' to perform and try to attain their objectives, while low motivation is expected to be detrimental to the attainment of the objectives. In either case of high and low motivation, it can be expected that a situation may arise where the user might try to game the system (cheat) in order to attain their objective *because* their motivation is high. In systems such as CAT (Computerized Adaptive Testing), the motivation of the users using the system can be considered high – they need to use the system effectively to score better. If the user chooses to abuse features within the CAT system to their advantage (deviating from the intended path of the objective), it can be extremely detrimental to the state of the system's measurement of the user. Thus motivation is a critical factor, having both advantages and disadvantages. Various studies have been conducted on how motivation varies with users and on methods to quantify, calculate and represent user motivation differently based on the domain. In time diary surveys, we propose that the motivation is different for respondents directly using the system (in SAM) and for interviewers using the system (in IAM). Interviewers are given the job to conduct the interview, and are thus assumed to possess high motivation to complete the interview. We also assume that the interviewers would not try to game the system and that they always try to record the information provided by the respondent as accurately as possible. Respondent's on the other hand, are seen as low motivation users. Their expectation from the survey is minimal, usually limited to a small financial reward. Since there is no relation between the quality of the data and the reward obtained, a respondent may resolve to providing responses that are easier to report than true, like saying they slept the whole day instead of listing out their individual activities. However, we also place some emphasis on the fact that when a respondent agrees to a survey, they have the minimum amount of motivation to do the same. During the course of the survey, this motivation might increase and allow them to successfully complete the interview or conversely game the system into completing the interview for them (by meeting the minimum required conditions for completion), or their motivation might decrease leading to a decrease in the data quality and subsequently resulting in a break-off. Thus measuring the motivation of the user is one of the issues that knowledge engineering has to deal with. Interviewers conducting interviews use intonation, speech speed and other verbal cues to both recognize the respondent's motivation and to guide the respondent to finishing the interview. However, they themselves are mostly unable to articulate all the rules or reasoning they use for this forcing us to propose alternate relatable methods to do the same within our framework.

The next aspect of the user is their purpose in using the instrument. This measure is almost identical for both interviewers and respondents – their purpose is to record the responses with the instrument. The difference occurs in how the response is obtained – the interviewers need to extract it from the respondent through conversation, while the respondent has to extract it from their memory and articulate it. While the process of extracting the information from memory is beyond the scope of this work, we pay attention to two proven methods of recollection: backtracking and visualization. For knowledge engineering, the purpose of the user defines what sort of knowledge must be made accessible to the corresponding mechanisms.

The final distinction between users comes with their knowledge of how to use the system. A user familiar with both the purpose of the survey (interview), and of their recollection of the responses are now faced with the task of representing their responses in the manner required by the instrument. Any shortcoming in understanding the survey or with their recollection will bring about a similar shortcoming when it comes to using the instrument on top of the problems faced with using the instrument. In the conventional scenario of being assisted by the interviewer, these shortcoming are addressed by the interviewer in possibly two ways:

1. They ensure that the respondent understands the purpose of the survey interview at the beginning of the interview completely, or
2. They provide sufficient information to the respondent to start the survey and then use a step-by-step approach in helping them understand the purpose of the survey by going through its requirements.

Also, in case of the interviewer-assisted interviews, the respondent is isolated from the instrument and the interviewers are usually trained beforehand on using the instrument (or the instrument is modified to fit within the understanding of the interviewer). Thus when the respondent uses the instrument directly, the framework needs to pay special attention to the increased amount of cognitive load now on the respondent and the knowledge engineering mechanism must identify and quantify it too.

Thus when tasked with modeling the interview, the knowledge engineering mechanisms must handle the ways to identify, quantify and use the user’s characteristics to drive the survey. Another obstacle that comes in view here is the source to obtain this information without resorting to more advanced technology needs (like face scanner, eye trackers etc.,) since that would counter the advantages of freedom and accessibility provided by using the web. The user responses and the paradata recorded during the interview process (and that is historically available) is the primary data source for the knowledge engineering mechanisms and it must be fitted for the purpose. While directly correlating data for the user's characteristics is and may not be available, the framework must use indirect means to achieve the same for the user (interviewer and respondent) and this is the objective of the knowledge engineering mechanisms.

3.4.1.2.3 Knowledge Engineering – The Solutions Proposed

As stated in Section 3.4.1.1, the knowledge engineering mechanisms aim to model the interactions that happen within the system. We break down the modeling process into the types of users first with the integrated approach taking priority. Thus we have two types of users to model – (1) the respondents and (2) the interviewers. The characteristics that we are interested in modeling are their (1) motivation, (2) their purpose, and (3) their knowledge about how to use the system. Once the modeling is accomplished, it can be used by interaction mechanisms to improve the user's experience while using the survey and thus bring about a better survey. The interaction mechanisms may themselves further require more knowledge engineering mechanisms to source their data and this will be discussed later.

3.4.2 Interaction Mechanisms

The interaction mechanisms are those mechanisms that help translate the agent's decisions into user interactions. The interaction mechanisms are considered as a separate problem to handle the different rules of surveys that we encounter as described in Chapter 1. This allows the framework to bring about sufficient flexibility to be oblivious to the type and method of passing data through the system as each mechanism (and module) can act independently on choice of data to use, process and output.

3.4.2.1 **Understanding Interaction Mechanisms**

At this point, these mechanisms are working to improve the following characteristics of surveys:

1. Make surveys faster

    This involves allowing better data entry in terms of interviewer-assisted mode and smoother data entry in self-administered mode. Faster data entry for self-administered mode may prove more detrimental than useful since it could lead to biasing effects.

2. Generate better data in terms of quality (response quality)

    This involves improving the quality of data obtained through the interview in terms of completeness and reducing errors. Completeness refers to minimizing instances of memory gaps and increasing recall when needed.

3. Prevent break-off

    This involves preventing the respondents from quitting once an interview has started. Unfortunately identifying and preventing break-offs is a complex process and as such, the framework alternatively includes keeping the interview interesting and the respondent well informed as the primary ways to accomplish this.

3.4.2.2 Interaction Mechanisms – The Environment

Interactions between the user and the instrument are tantamount to the successful understanding and usage of the instrument by the user. During the course of a time diary survey, the interviewer and the respondent undergo different interactions with the instrument. They have different expectations from the instrument and also expect different behaviors from it. Thus an interaction that works for the interviewer might *not* work the same for the respondent and vice versa. A simple example for this is providing the interviewer with predictive lists. While this would be a feature appreciated by interviewers for the time it saves them typing the data, when delivered to respondents, it becomes susceptible to introducing satisficing thus becoming a negative feature.

The interviewer expects an instrument to serve as the recording tool for the information they elicit from the respondent. As a recording tool, it can be expected that there would be consistency in how it looks and behaves and must be geared towards entering and submitting information well. Additional features that transform the data into formats that the interviewer can use during the survey can also improve the interviewers' acceptance of the instrument. Table 4 lists the expectation of the instrument behavior and its tasks for an interviewer and a respondent.

Expectation	Interviewer	Respondent
Expectation of instrument behavior	Behave as a recording tool	Behave as an interviewer
Expected tasks that the instrument must perform	Record data, allow for fast recording, consistency and error checks	Provide information regarding how to use the instrument, guide through steps required to complete interview, assist in identifying and handling errors and providing

		information on progress of interview and its completion

Table 4 Interviewer versus respondent expectation of instrument

The respondent however has no intermediary when directly using the instrument. Thus they expect the instrument to provide them with all the required information to start, proceed and complete the survey. For respondents thus, the instrument is expected to behave like an interviewer and interact similarly – guiding them through the survey, getting them to provide their responses and record them in the instrument, assisting them in identifying and fixing errors and gracefully exiting the survey. Thus, for the respondent, the instrument must be geared towards reducing the cognitive load requirement through the interview process.

3.4.2.3 **Interaction Mechanisms – The Issues Faced**

The issues faced in designing Interaction Mechanisms primarily arise from attempting to describe what interaction is needed for the user (respondent and interviewer) and how to deliver the interaction in a way appropriate for the user. As mentioned earlier, what works for the interviewer may not simply work for respondents, but might also be detrimental. Each Interaction Mechanism can influence the user in varying degrees and must fit within the design of the overall framework. If we were to consider the two users differently we would beat the purpose of integration. Thus the two users must be differentiated and handled with only as much separation as needed.

3.4.2.4 **Interaction Mechanisms –The Solutions Proposed**

Time diary surveys are essentially conversational surveys like the work of Kite, 2007. However, unlike their Event History Calendars where the memory recollection (and its handling) is of the highest importance, our work focuses more on the interactivity/usability of the system for both interviewers and respondents. This focus is motivated by the many challenges posed by the survey domain as mentioned in Chapter 1.

The mechanisms to bring about this interactivity and usability are primarily:

1. Probing

   Probing mechanisms are those by which the agent exacts information from the respondent at specific circumstances. These circumstances are varied, such as when the user is in the process of creating an activity or filling a particular value, at a point when the system identifies that the user needs to pay attention to a particular piece of information etc. General probing mechanisms in intelligent systems as described in Chapter 2 are not usable directly in our framework since frequently spamming the user with dialogs/messages can unintentionally cause break off due to the reduced motivation of survey respondents (unlike users mentioned in Chapter 2).

2. Autocomplete

   Autocomplete mechanisms are primarily a feature of the interviewer-assisted mode wherein expert and novice interviewers can increase data entry speeds by not having to type complete content. The autocomplete mechanisms can also assist in providing live de-identification support by parsing the verbatim as it’s typed. By coupling this with an online KE mechanism, the framework could potentially bring about a substantial increase in the data entry speeds by providing the interviewers with predictive content.

3. Using precodes

   Precodes are a mechanism by which content for certain fields in the instrument are displayed in a much more accessible format (for example clickable boxes). The content within these *precode lists* can be live if needed (in case of self-administered) changing as per the respondent models or pre-defined with expert advice. This allows bringing about a certain amount of expert knowledge into the system further increasing his utility.

4. GUI design

   The general GUI design is also a core part handled by the framework since the GUI itself has implications for surveys. The UI must be easy to use yet not bring about biasing or influence the user to prefer one response over another due to simplicity or ease. Furthermore, the UI must be adaptable to different delivery mechanisms (such as PC screens, mobiles etc.) and be flexible for switching between interviewer-assisted and self-administered modes. By keeping the design of the instrument as part of the framework, we are able to address concerns about usability and can add utility to the framework by improving the design to suit the target user.

5. Software Assistance

   The software referred to here is the instrument that is visible to the user – hence it's the part of the GUI and the agent interactions that are available to the user. Software assistance hence refers to those mechanisms by which the process of learning to use the instrument and the general use of the instrument itself is *assisted.* This assistance can be provided using relevant help (which can be coupled with online/offline KE to support personalization), resources to help respondents in self-administered mode to easily use the instrument. Hence this component is vital to keeping the learning curve of the instrument as smooth as possible for both interviewers and respondents. This effectively helps in handling the issue of lack of motivation and user experience frequently encountered in the survey domain as illustrated in Chapter 2.

6. Attention capture

The attention capture mechanisms are intended, as they imply, to capture and keep the user's attention to the task at hand (the interview). This may be accomplished by coupling this with the relevant KE modules and other delivery methods such as Probing to divert or direct the user's attention to particular event such as missing a value (by assisting recall) or initiating interactions with the user when they deviate or have been facing difficulty moving through the interview (Marinilli, 2003, Shneiderman & Plaisant, 2005). By handling the mechanisms that modify user attention separately, sufficient separation can be bought about between interviewer-assisted and self-administered modes. The aforementioned mechanisms are further summarized using Table 5 as to what they are envisioned to accomplish.

	Faster survey	**Higher quality**	**Prevent break-off**
Probing	YES	YES	YES
Autocomplete	YES	-	-
Precoding	YES	-	-

Software Assistance	-	YES	YES
Attention Capture	-	YES	YES
GUI Design	YES	YES	-

Table 5 Interaction Engineering Mechanisms and the framework characteristics they attempt to fulfil

3.5 The integration of the framework

The design of the framework is intended to handle both respondents and interviewers within the same architecture without requiring to have different instrument/design intended to cater them separately. This, as explained previously, is partly due to the differences in how respondents and interviewers would envision the survey to be from their own perspective. This difference in how the user interacts with the instrument must be reciprocated by the instrument also. A simple example of this is how many predictions can be provided to the user. An interviewer may be shown the top 5 (or more) predictions since it can be safely assumed that given their high motivation in doing the interview, they would not attempt to satisfice or be overwhelmed by the predictions. However, for a respondent, providing the 5 predictions might be more detrimental than useful and it might be a better idea to show them a reduced set (of maybe 2) and in an appropriate manner so as to provide assistance, without overwhelming them. Figure 9 illustrates the process for generating predictions for interviewers and respondents.

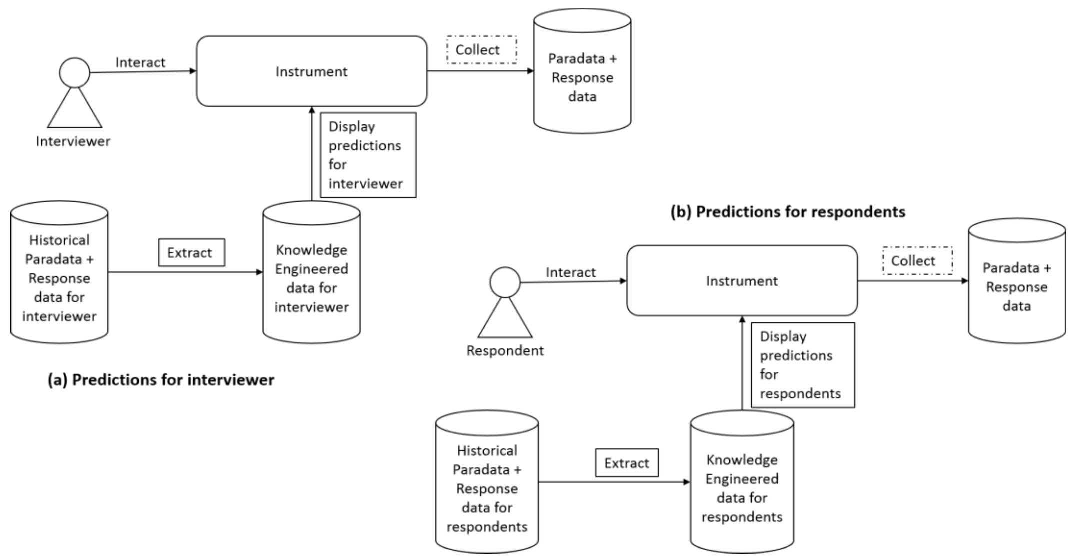

Figure 9 Process for generating and displaying predictions for respondents and interviewers

Figure 9 illustrates the basic process for collecting data, generating and displayed the prediction information to the two types of users. The very similar flow employed for the two users obfuscates the differences for the two users when collecting the data, generating and displaying the predictions to the user. The solid border actions; Extract and Display predictions for interviewer/respondents); denotes the actions that need to change or adapt to handle the two user types. The dotted border action (collect) need only be the same for both the users so as to collect and record the data generated during a survey. Combining the processes for both users in such a way that the difference in operation exists only where needed enables the framework to integrate the differences required for both under one umbrella. This is achieved by breaking the framework into component mechanisms that perform certain unique tasks. Thus a Mechanism (pl. Mechanisms) is defined as a framework component that performs a task or a set of tasks that help in bringing about the execution of the required function. The only rule that the framework uses for declaring a Mechanism is that it performs only one core task. If there are multiple tasks that can be broken into more than one core task, then each of the core task needs to be performed by a different Mechanism. This helps the framework achieve separation of concerns together with simplifying development and easing maintenance.

This integration can be further exemplified using Table 6. With this table we present how the mechanisms can be switched or modified to handle respondents and interviewers separately while still maintaining the same architecture and accessing the data obtained in a user-agnostic manner. Thus individual mechanisms (or chains of mechanisms) can be modified at different stages as needed.

Mechanism	**Sub**	**Data source/type**	**SAM**	**IAM**
Prediction Mechanisms	Knowledge Engineering Mechanism	Response data, Paradata - historical	Top 2 predictions	Top 5 predictions
	Interaction Mechanism	Knowledge Engineered data	Display detailed information invoked by user inaction	Display as a list invoked by ‘Activity Creation’
Probing Mechanisms	Interaction Mechanism	Response data, Paradata - current	Explanatory probe invoked by probe requirement	Direct probes invoked by probe requirement
Autocomplete Mechanisms	Interaction Mechanism	Auxiliary data	Searchable, delay invoked and NLP involved	Searchable, immediate

Table 6 Illustrating the differences in SAM and IAM for the mechanisms

3.6 A more grounded view of the framework

The framework can thus be represented at an intermediate level as shown in **Error! Reference source not found.**10. Here the two instrument blocks (1 and 2) represent the instrument in the two modes; interviewer-assisted and self-administered modes respectively. The interviewer-assisted mode collects data into representation datasets D1 and D3. D1 stores the direct survey data collected from the instrument (such as activities) and may not be usable directly if the data has not been de-identified. This is extremely important in surveys since any identifying information from the data exposed to the framework must be eliminated as required in surveys. Thus the framework makes room for this by applying the required processing (online or offline) on the collected survey data before using it through the dataset D2.

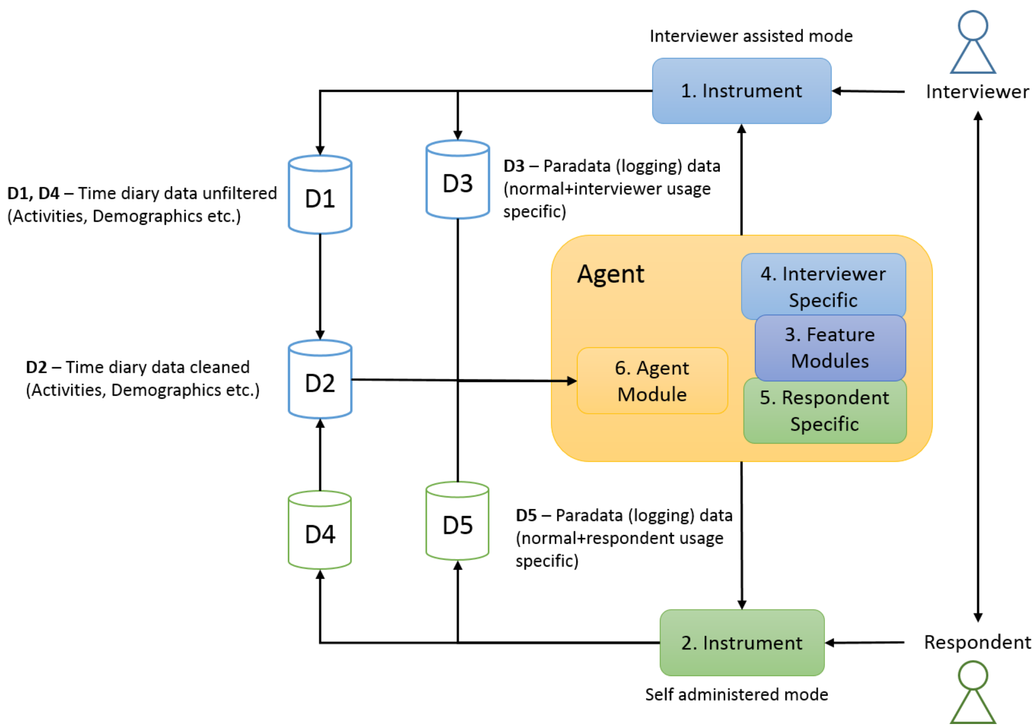

Figure 10 Intermediate level representation of the integrated framework

The response data thus obtained from either the interviewer-assisted or self-administered mode will be available for KE through D2. D3 and D5 represent the paradata and usage tracking logs from the two modes. This allows the framework to address the actions of interviewer assisted and self-administered mode differently. The interviewer assisted mode can be taken as a form of supervised learning for the self-administered mode providing important information for comparing and compensating between interviews conducted under the two modes. This essentially means that the interviewer assisted interviews can be modeled and used as baseline for the self-administered mode. Thus the system can bring about a learning process wherein, it learns from the interviewer and can translate the actions to the self-administered mode to bring about the advantages of F2F interviews as described in Chapter 2.

3.6.1 Framework Organization

In this section, we describe how the framework functions to achieve its purposes. The offline and design-time Knowledge Engineering Mechanisms (KEM) are executed ahead of deployment to generate the requisite data for the working of the other dependent mechanisms in the instrument. This allows for computationally expensive operations to be completed ahead of time, so that the data may be accessible when the system goes online. A Knowledge Engineering Mechanism can access the data that it is dependent on for performing a particular function and is invoked and executed when appropriate. On being invoked, the KEM executes based on the data at the point of execution to either output the required data (if processing) or creates the required intermediate data that may be used by down the line KEM or IxM (Interaction Mechanism). When a KEM is invoked based on some data outputted by another KEM, we essentially create a chain of action reaction. IxM on the other hand, are invoked based on KEMs. If the IxM for the user types are distinctly different, then there can exist two forms of the IxM for the two user types. The effect of IxM is usually on the respondent and they provide an implicit feedback through their next actions. These can be picked by the KEM to continue the process for adaptation and learning. An agent entity is created in the back end for each survey session being delivered. The agent's task in the back end is to control the flow of the processes when needed and to host the KEM usable for adapting and learning the users with time (currently unimplemented).

3.6.2 Putting it all together

When the framework is deployed and the necessary data (user information, auxiliary data) is available, it becomes ready for use. The KEM and IxM are capable of receiving inputs from multiple sources and directing their outputs accordingly. By controlling the execution of KEMs, the flow of the data through the system is also controlled and is also responsible for controlling downstream actions. For example, the detection of the start of a new activity invokes the appropriate dependent mechanisms, such as the Prediction KEM. This also separates the flow from the mechanisms directly as the mechanisms can be modified later on without breaking the flow. The detection of the different events is based on the interaction paradata and the response data collected by the data capture systems.

When a user begins using the instrument, depending on the user type, some IxM become inactive, some modify themselves accordingly while others would remain the same. With the start of the interview, IxMs that need to be invoked by this event get activated. Respondent-directed IxM such as for wizard type assistance (future work) that are only required for respondents can thus be uniquely activated under necessary conditions alone. When the user begins interacting with the instrument, the IxM send their output data to the be persisted in the database working in tandem with its backend Data Recorder KEM. A view of these flows is illustrated in **Error! Reference source not found.**11.

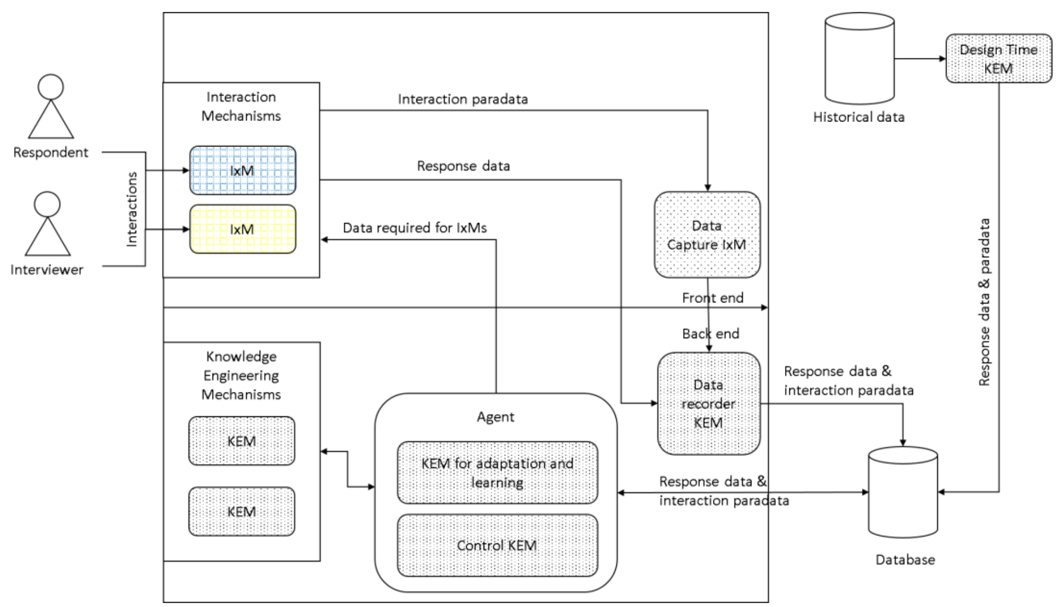

Figure 11 Flow illustration within the framework

During the process when the data is transmitted, converted and persisted into the database, the agent can use the data to control the KEMs. The KEMs are invoked at the backend when the appropriate conditions are met (e.g. new activity data incoming, before persisting the data etc.). The accessibility of the database by KEM and IxM is hidden by the presence of the Agent, but in essence they have access to the database through the Agent which can control where the database is and how it is to be accessed. This also allows the introduction of a deeper logging mechanism at the agent for future use. Once a KEM has finished its current execution cycle, the data (if any) is fed back into the database using the agent. This can begin the execution of the next IxM and/or KEM that might require the data. Note that our Agent module here serves as a controlling system for the interview instance. It is essentially a shell through which we can keep track of how and what data flows through the system. This would become an integral part when integration of the two modes is taken into account and when adaptive learning is added into the system.

When the user interacts with the instrument, all the interactions are logged using a Data Capture IxM that forwards the required data (and context such as time) to the backend for persistence. When the user interacts with the IxMs, they continue to provide the appropriate events for the system both due to action or inaction allowing the flow to restart as needed. The integration of the two modes allows the framework to envision future mechanisms that can use data from one mode to power mechanisms from the other mode. This could help the system to 'learn to behave like an interviewer'.

3.6.3 Summary

Thus we build our framework from ground up using a highly abstracted view of the environment of time diary surveys. We deal with how interactions are different among the users and how the instrument can interact with the user within the rules of the survey domain. The different mechanisms proposed and later built leverages existing technologies which can later be expanded with more advanced technologies to handle more difficult problems. The design and structure of the integrated framework takes into account the fact that not every problem may be solvable at one go (or at the onset) when it comes to human computer interactions and keeps the mechanisms loosely coupled but highly cohesive to enable a cumulative way to approach the problem.

Chapter 4: Implementation

4.1 Introduction

As part of the implementation of the framework described in Chapter 3, we designed and developed a software prototype instrument based off of a planned multi-phase implementation. The instrument is called the 'Web ATUS' instrument and is designed to be used primarily by interviewers to conduct interviews with parts of the framework working to assist the interviewer. However, our design of this instrument considered the dual-role of the instrument in later stages as being used for administering both interviewer-assisted and self-administered mode. This is accomplished by designing the instrument as the *current interviewer-assisted* implementation of the integrated multi-mode framework. One of the key design features that enables this is within the design of the database system where the data is stored. By abstracting the way that the knowledge for and within the system among the different mechanisms is stored, we create avenues for a natural way to incorporate multi-mode features into the system – we explain more on this in Section 4.2.4. In this chapter we present our prototype instrument, its functionalities, workflows, and how future components (or mechanisms) can be integrated into the implementation architecture.

4.1.1 Study phases

The implementation and its subsequent testing are planned in three phases to enable a more feedback based development approach where the data from the preceding phase is analyzed to improve the next phase. This is a necessary part of the development cycle for this instrument due to the high amounts of unknowns when it comes to how users interact with the system. Since the implementation does not attempt to pigeonhole the user into using specific methods to enter and view the data, the implementation is designed to evolve as more information (knowledge) is attained during each testing phase. A brief overview of the three phases are provided below:

4.1.2 Phase 01 [June 2014 – July 2015]

Phase 01 is aimed at testing the viability of an instrument implementation of the framework described in Chapter 3. The primary focus for this phase deals with knowledge engineering, instrument design suitable for delivering time diary surveys with administrative capabilities and at testing the suitability of the interaction mechanisms.

The implementation completed Phase 01 design and testing during the months of June and July, 2015. It must be noted that the design and development of the implementation has been running throughout most of 2015 and the later parts of 2014, while the testing was conducted beginning June 2015. Four students from UNL's Bureau of Sociological Research (BOSR) were recruited and trained to play the role of the interviewer as part of the interviewer-assisted mode. The students had had some familiarity with general interviewing techniques and were quickly able to grasp the concept of conducting time diary surveys over the telephone using our instrument for recording data. Forty-eight respondents were recruited by the means of advertisements and posters from in and around Lincoln, Nebraska. Equal number of male and female respondents of the three age groups were selected and interviews were setup with them beforehand. A screen video capture software named Camtasia® was used on the computers used by the interviewers to record both the audio and video of the instrument during the interview.

4.1.3 Phase 02

Phase 02 of our studies is intended to serve as the means to verify our changes to the instrument following analysis of the data from Phase 01. After analyzing Phase 01 data, it was observed that the interviewers did not use the recommendations provided to them as intended, i.e. they did not click on the recommendations. However, our analysis of the interviewer videos led us to believe that the interviewers might be using the recommendations for visual cues. Thus certain design changes were incorporated on both the GUI and the data collected. The testing of Phase 02 began in November, 2015 and continued well into March 2016. At the time of writing this thesis, Phase 02 was only partially completed.

4.1.4 Phase 03

Planning for the objectives of Phase 03 is still in progress and is expected to be confirmed once the data from Phase 02 is obtained. The current tentative objective is to incorporate the design with self-administering mechanisms and increase the learning capability of the instrument.

4.2 Instrument Prototype

4.2.1 Introduction

The instrument is implemented as a client – server architecture model. The client is a web application executing as web pages delivered to the user's browser over HTTP. The client in our implementation is a rich client; many computations limited to the client side are performed on the client's browser itself and the client has almost direct access to their own data. This is further supplemented using a RESTful (Representational State Transfer) server application that supports distribution of load. A block diagram of the implementation architecture is shown in Figure 12.

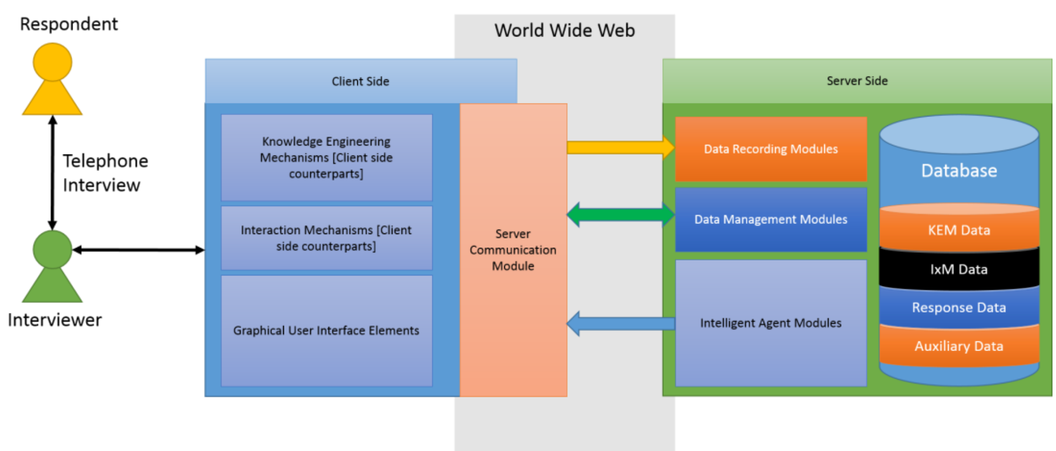

Figure 12 Block diagram representation of the implementation architecture

4.2.2 Server design

The server (backend) for the implementation is written in Java (JDK 1.8) and uses the Representational State Transfer (REST) architecture to communicate with the client over Hyper Text Transfer Protocol (HTTP). Using REST allows the server to deal with networked applications wherein multiple clients connecting to the system can deal with their data individually. It supports scalability and maintainability and thus allows our server to be both scalable and maintainable. This programming model also supports modularity. The server application was hosted on the Intelligent Agents and Multiagent Systems (IAMAS) lab's server at iamas.unl.edu allowing for access over the Internet. A simple authentication protocol was used to verify the interviewers since authentication was not a priority in the current phase. However, the application is designed to be adaptable to any authentication model like OAUTH etc., at later stages without affecting any other modules including user management. The server serves two core purposes:

1. To handle the flow of data from the client, and
2. To provide the knowledge required by the client as and when required. This is performed by an agent at the server end.

4.2.3 Client design

The client application was written as a web application executable with any modern web browser. It is a rich client model where most independent client actions are performed at the client-side as opposed to the server-side as is with conventional server-client models. This alleviates the load on the server and distributes the load at the client level itself, making it scalable. The client web application was written in HTML5 with CSS4 and uses JavaScript extensively. The application is writing in the Model View ViewModel (MVVM) architecture pattern which allows for separation of the design elements (views) from the data elements (models). This further allows for modularity which is necessary when considering the fact that certain Interaction Mechanisms may need to be turned on or off depending on the user (respondent or interviewer). The Interaction Mechanisms are implemented using the client side code and are further explained in Section 4.2.6.2.

The client makes use of external JavaScript libraries to handle many of core functionalities with separate modular JavaScript code for the application's use. The external libraries used are:

1. jQuery

   The application uses both the core jQuery script and the jQueryUI script for design related tasks like dynamic controls (dropdowns, autocompletes etc). The version of jQuery used is 2.1.1.

2. LINQ (Language Integrated Query)

   LINQ is an extension to a programming language by the addition of query expressions like SQL statements to many enumerable types of data like arrays, collections etc. It is in fluent-style where commands can be fluently chained to one another to be almost read like English. It allows for manipulation of lists and arrays and is a significant contributor to reduce boilerplate code (e.g. for loops to find maximums, loops to sort etc.). The JavaScript port for LINQ is called linq.js.

3. Knockout

KnockoutJS is a library that brings about the MVVM design pattern to HTML/JavaScript. As mentioned earlier, MVVM allows for separation of the GUI elements from the data elements allowing for modular design.

4. Timeline

    The timeline.js script file allows the use of a horizontal timeline where events can be represented against a chronological timeline that supports zooming in and out, panning, selection and moving events from one point of time to another on the timeline. It is a bootstrap for Google's visualization engine and hence is powered internally by Google code.

5. Miscellaneous data manipulation libraries

    We also make use of two data manipulation libraries called string.js and Date.js that implements many core string and date manipulation methods that are otherwise unavailable in JavaScript thus further reducing boilerplate code and allowing us to focus on the primary application code.

All the above libraries support a minified version of their code that usually reduces their file size by around 80% to 90% addressing the concern if using the libraries would increase page load times and such. Increased page load times has been known to have a detrimental effect on the user's interest in using applications.

All the application code is written with an *atus* prefix and is modular in design with each script handling one core functionality of the application. The application scripts are:

1. atus-resources.js: This script file contains all the static and constant data for the application such as the strings used to denote activity types, fields and context information. These are generally known as magic strings (strings that magically have a purpose) and in good design are generally shunned from usage within application logic code. This script also contains the different messages that are shown to the users. Abstracting such magic strings to a resource file allows for localization and significantly reduces errors introduced by the use of magic strings (e.g. having ‘ActivityNme’ instead of ‘ActivityName’ (the magic string) is an error introduced due to a typo in the code). Such errors usually result in logical errors which are hard to pinpoint and correct.

2. atus-server-com.js: This script file contains all the methods that allow for communication with the server. All the methods are bootstrapped to a SERVER object that can be accessed from anywhere within the other application scripts. This is a part of the modularity and the maintainability of the code. Having this abstraction allows the communication logic to be pushed to this script preventing errors and having a central location for logging all communication with the server. All communications to the server occur as Asynchronous JavaScript and XML (AJAX) post and get calls depending on if it is an update or read call to the server.

3. atus-internal-vms.js: As the application uses the MVVM pattern, this script is where the internal ViewModels used by the application resides. Currently this is

limited to the _precodeVM class that handles the use of the precoding mechanism.

4. atus-session.js: This script handles all the session related actions for the application such as initialization (communicates that the client is requesting a new session), loading the assets used in the application (the activity names and codes, the who and where names and codes and loads the user's information to be used in the session.

5. atus-paradata-tracker: As the name suggests, this script file tracks all the user button clicks, keystrokes and other paradata information such as field entry and exit times independent of the application itself.

6. atus-prompt.js: This script file is the modular script for managing the prompts delivered to the user via the GUI. These prompts currently recommend the TOP 5 next activities based on the two prediction methods (obtained from the server) to the user (Phase 01). It handles the creation, display and the removal of the prompts.

7. atus-activity.js: This script file defines the Activity ViewModel class called _activityViewModel. This class deals with managing the display, change and features that accompany the management of an activity within the client. This includes the logic for switching the context information based on the activity (loaded during session initialization), determining the validity of the currently set data etc.

8. atus-dialogs.js: This script file defines the _dialogVM ViewModel class for creating and displaying the dialogs windows associated with overlapping activities. This can be extended to do the same for other types of dialogs that may be needed at later stages.

9. atus-overlap.js: This script file contains the ViewModels required for storing the overlap information of activities that can be used by the _dialogVM to display the same to the user.

10. atus-page.js: This script defines the core _pageViewModel that handles all the functionalities within the page such as managing the activities recorded by the interviewer, the logic for the interview state, and the logic for deciding when to display the different prompts and dialogs. It uses the other ViewModels within it to connect the different activities and their information to the instrument.

11. atus-run.js: This script contains the primary initialization code that begins initializing and loading the instrument's engine.

12. atus-engine.js: This is the bootstrap script for the instrument that is initiated by the atus-run script (when the page is loaded) and contains the _engineViewModel. The engine ViewModel creates the page, begins session initialization, manages the communication between the page and the timeline, manages the resources and binds the different modules together.

The script files are all in pure JavaScript and thus can be run from any modern browser such as Internet Explorer 10+, Microsoft Edge 12+, Mozilla Firefox, Google

Chrome, and Opera. The limiting factor for the browser is the ability to handle AJAX calls for REST and render HTML5 correctly for rendering the timeline and the instrument's GUI. Another limiting factor is the screen resolution, minimum required 1280 x 768, since that is the minimum required screen real estate for displaying all the panels and controls correctly without causing overlaps and breaks in design due to lack of space.

4.2.4 Database design

The database for the prototype web ATUS instrument was designed to be relational and uses MySQL as the SQL server. The database is designed to be extensible based on the modules implemented thus favoring the addition and modification of both Knowledge Engineering mechanisms and Interaction Mechanisms as and when needed. It is principled to ***separate*** data based on its use and is ***modular*** in most instances. The use of relationships then allows the data to be related to each other creating the knowledge that the system uses and creates.

There are currently 33 tables in the database (inclusive of one extra table added for Phase 02). Three of the tables (mappingstbl, versiontbl and interfacetbl) currently act as placeholders for the integration of future modules and for localization support if needed.

4.2.4.1 **Data Separation**

The database stores the knowledge required for the Knowledge Engineering Mechanisms and Interaction Mechanisms separately to allow for separation of concern. Data separation allows the system to be flexible in its extension wherein the Knowledge

Engineering Mechanisms and the Interaction Mechanisms can be modified and extended almost independently without requiring changes on each other. Data separation refers to *the practice of keeping code separated from the data it uses*. This typically involves a behavior in the code wherein, the code does not 'magically' know or use immutable values and time-variant values from within the programming logic. This kind of data is abstracted or separated into a distinct layer where these properties are stored which the code then uses to understand how to use the aforementioned data. Our implementation extends upon typical data separation wherein, the data for different modules (mechanisms) are also separated so as to provide the modules freedom to extend or change their data without severely affecting the working of other modules.

The Knowledge Engineering mechanisms related table structure is illustrated in Figure 13. Here, the conceptstbl, conceptactivitytbl, conceptwordstbl and conceptverbatimtbl store the required data that is needed for the Activity-Concept Translation Mechanism to perform basic language processing to attempt to convert verbatim responses to the coded activities within the system. This set can also be used by the future implementation of Natural Language Processing mechanisms to further ease the effort required by respondents in self-administered mode while filling out the activity information and related contextual information. The five associated stats suffixed tables – activityfieldstatstbl, wherefieldstatstbl, whofieldstatstbl, todstatstbl, and sequencestatstbl contains the knowledge engineered from both the design time knowledge engineering and the prediction mechanisms.

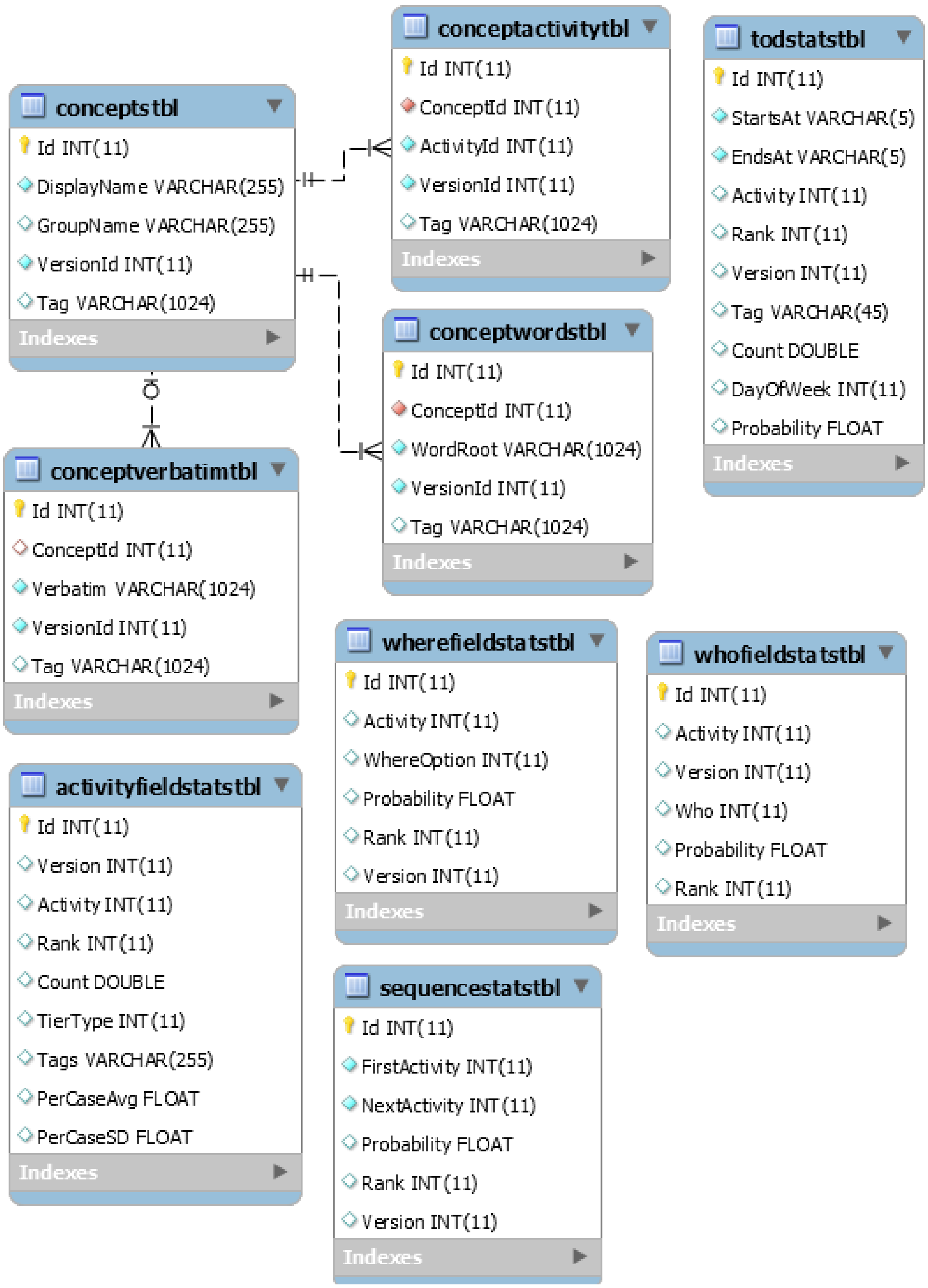

Figure 13 Knowledge Engineering Mechanisms related tables

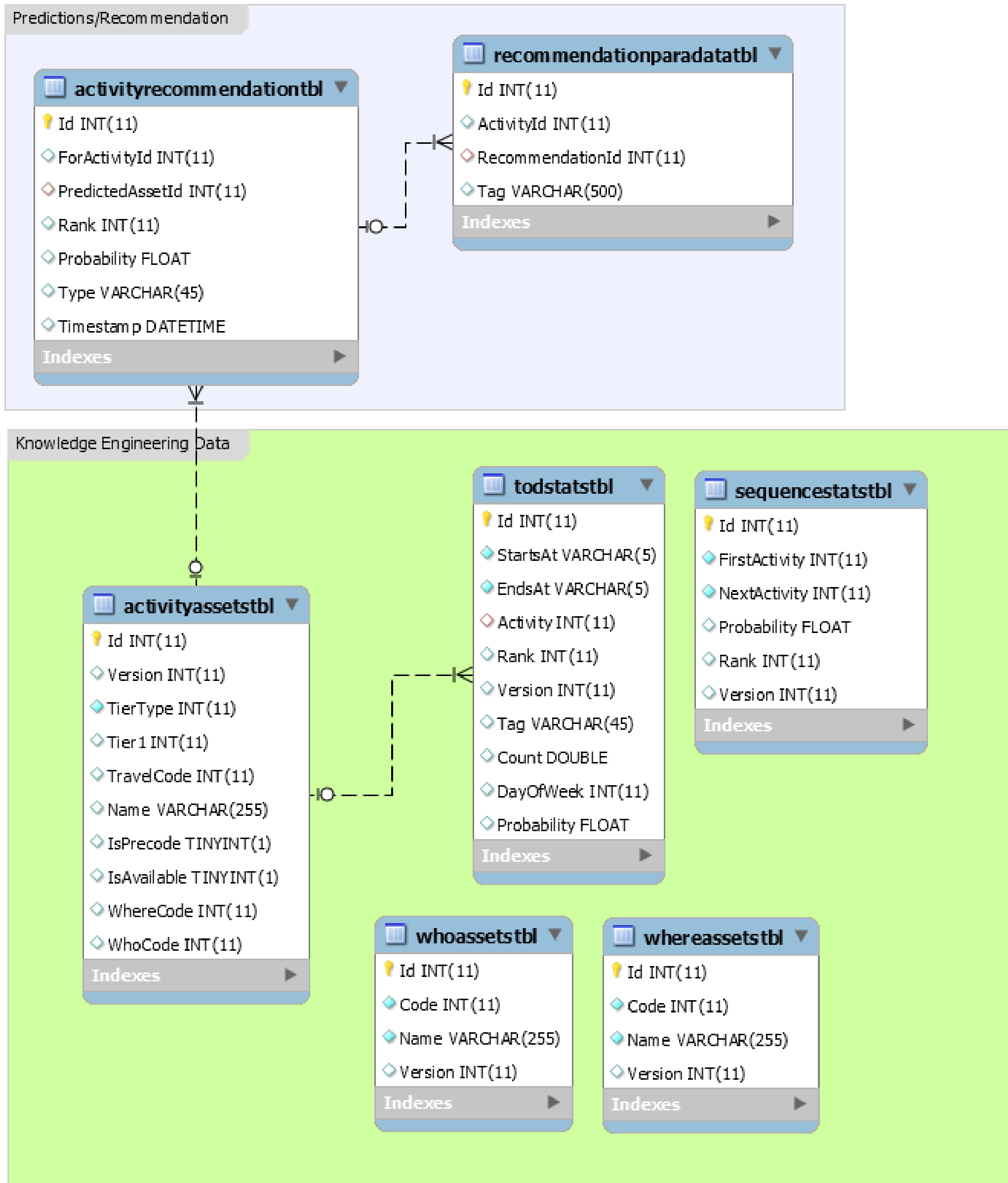

Figure 14 Interaction Mechanisms Tables (and associated Knowledge Engineering Mechanisms data used by Interaction Mechanisms)

Figure 14 illustrates the interaction mechanism tables (activityrecommendationtbl, and recommendationparadatatbl) that the Prediction Mechanism uses within itself. As is also shown, the Prediction Mechanism also uses the data from the Knowledge Engineering Mechanisms to make predictions and relates to the activities from the activityassetstbl. As an example of the application of the advantage of data separation, following the analysis of Phase 01 data, we made some changes to the Prediction Mechanism module to shift the predictions from a separate panel to within the precodes itself without any subsequent breaking or reworking of any other modules.

4.2.4.2 **Paradata Tracking**

The database also has a dedicated set of tables used to store the tracked paradata. This begins with the user's browser and system information (without any identifying information such as IP address) when they log in to the system and then the button clicks, field entry and exits, keystrokes as they use the application. The application also tracks the user's interactions with the timeline. The data from these are sent directly from the client to the corresponding REST methods on the server which persists them on the database – as mentioned earlier, RESTful methods provide almost a direct connection for the client to their data. The data tracked is illustrated in Figure 15 together with their relationships to the auxiliary and response data – so as to identify the context of the paradata. The activityswitchsequencetbl is a new table introduced in Phase 02 to track how the user switches between the activities directly. Figure 15 also shows the recommendationparadatatbl which is a table where the paradata regarding the usage of the Prediction Mechanism is stored. This is an example of the modularity that went into the design – this data is managed by the Prediction Mechanism itself thus allowing it have full control on it.

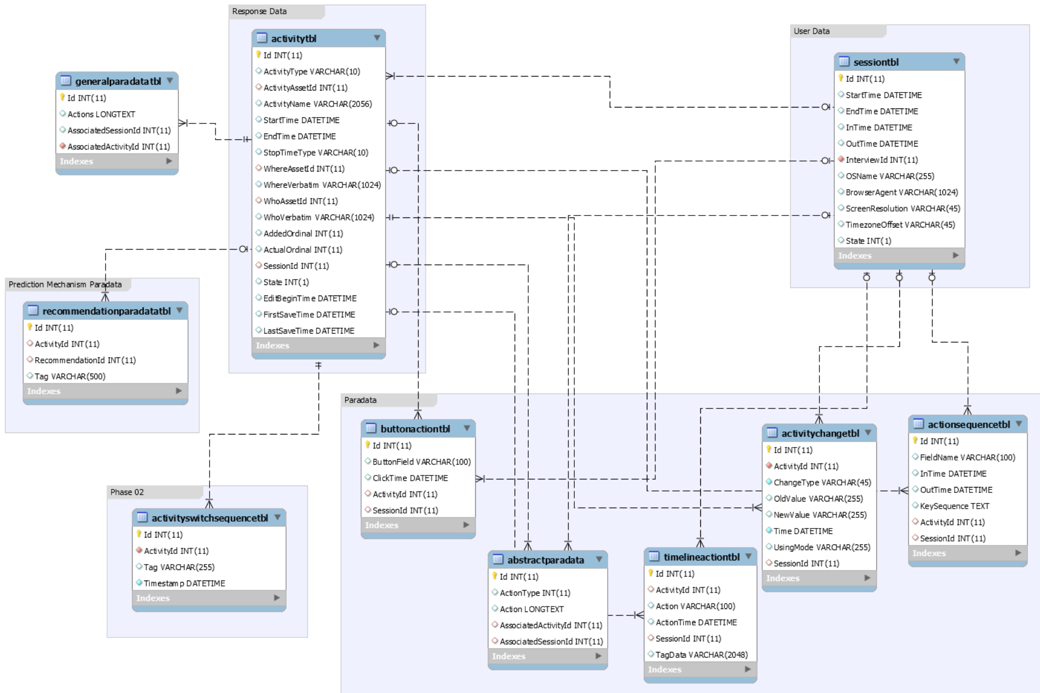

Figure 15 Tables associated and related to paradata tracking

4.2.4.3 **Response Data**

The user's response data such as the activities and their associated context information is stored in a separate set of tables with relationships defining their associations to the Knowledge Engineering Mechanisms data, the Auxiliary data, the tracked paradata and the user data. The tables associated with the user's response data are shown in Figure 16. Though the response data is associated with other data, it is not shown in the figure due to space constraints. The activitytbl stores the activity name as both the verbatim response recorded by the interviewer and the associated auxiliary data Id if the system could determine it using Knowledge Engineering Mechanisms. In a similar manner, the context information such as the Who and Where responses are stored as both the verbatim response and the associated system identified asset Id. This is an example of the post processing that is performed in real time on the data to enable it to be used by other mechanisms.

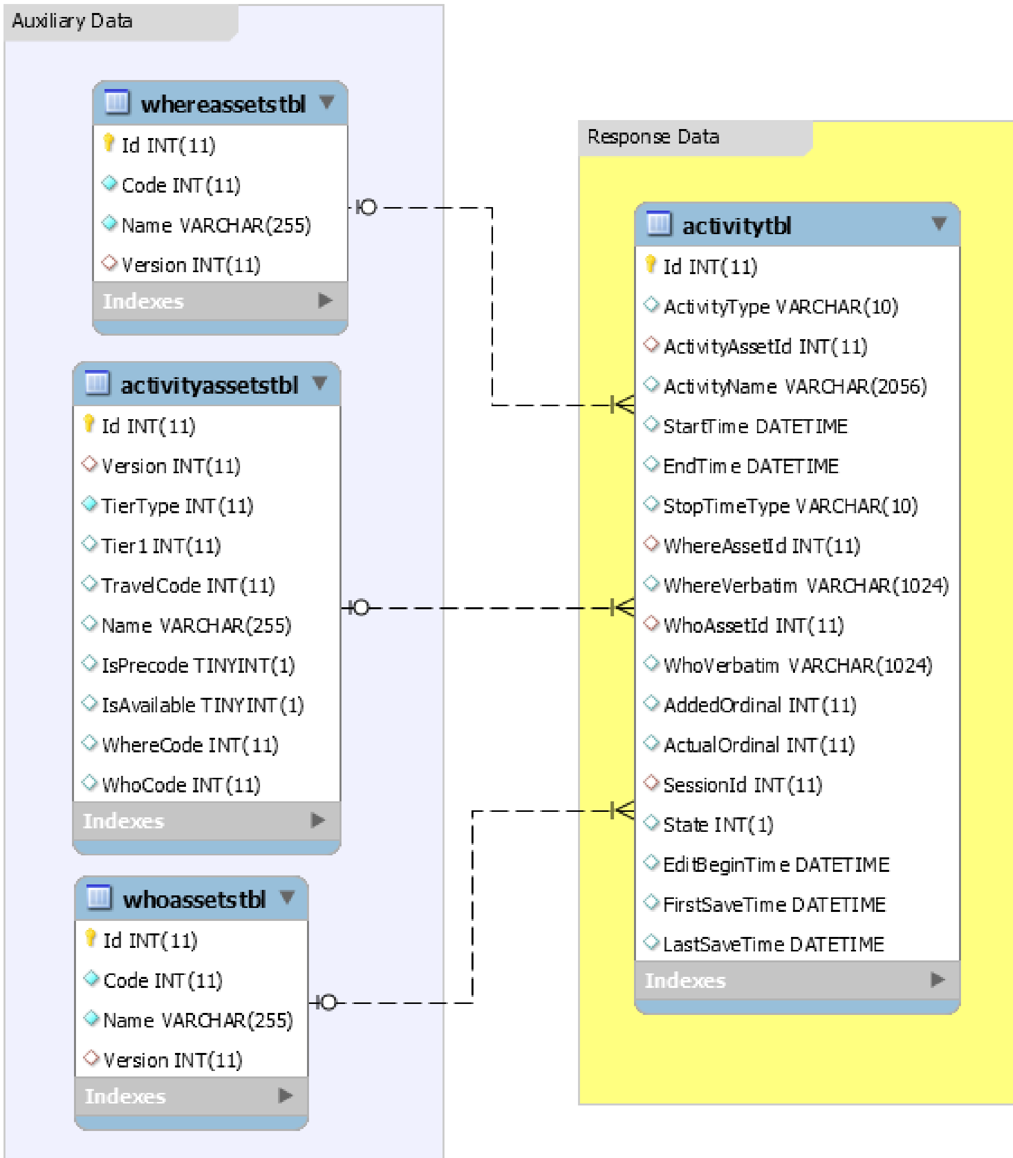

Figure 16 Table associated with Response data and the relationship to the Auxiliary data

4.2.4.4 **Auxiliary Data**

The auxiliary data in the system is the data that forms a part of the knowledge that is shared among multiple modules and mechanisms. This data currently consists of the assets (otherwise called identified) data concerning the activity, who and where information. These tables are shown in Figure 17. The auxiliary data was generated by the design time knowledge engineering mechanisms. This data is thus relatable to the domain knowledge that the system possesses.

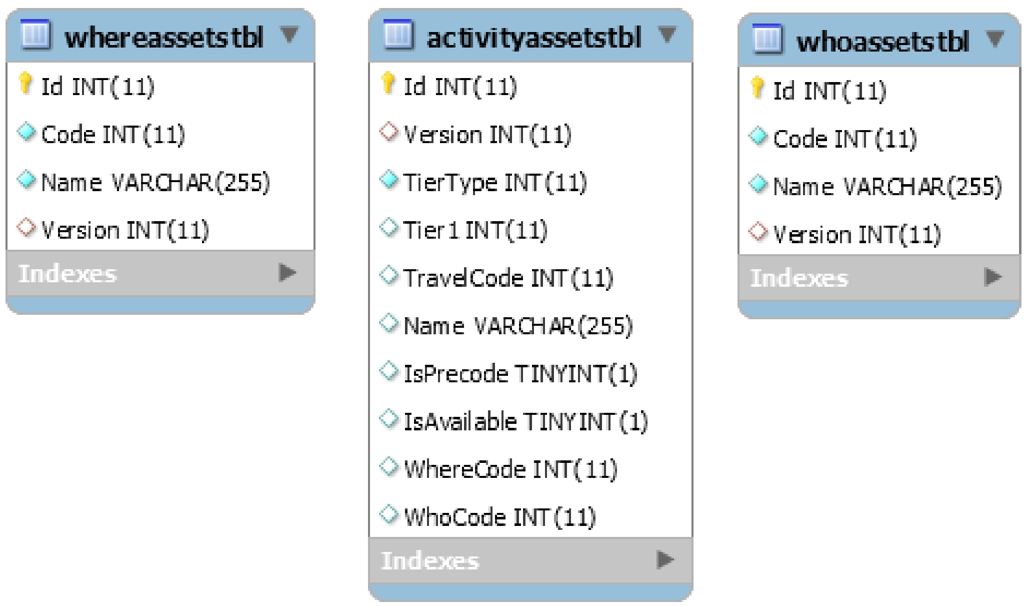

Figure 17 Tables associated with the Auxiliary data

A few rows from the activityassetstbl have been listed in Table 7 to provide an example of what this data entails. The columns Version, TierType, and IsAvailable are not shown since they do not play an important role in this implementation. The Tier1 columns refers to the validity of the activity as an identified activity – any value less than 0 implies that it is not a real activity and is a placeholder (eg. Refused and Don't know/Can't remember). The TravelCode indicates if the activity is a traveling activity, while the IsPrecode signifies if the activity should be displayed in the precode list. The WhereCode and WhoCode denotes if the values are optional (0), mandatorily required (1) or mandatorily not required (-1).

Id	Tier1	TravelCode	Name	IsPrecode	WhereCode	WhoCode
4	4	NULL	Personal care	1	-1	-1
9	9	NULL	Educational activities	1	0	0
10	10	NULL	Religious activities	1	0	0
12	12	NULL	Lawn care/backyard activities	0	0	0
13	13	NULL	Listening to music	1	0	0
14	14	NULL	Dancing and other performances	0	0	0
28	28	NULL	Reading	1	0	0
40	40	1	Traveling	1	0	0
46	46	NULL	Volunteer activities	0	0	0
52	52	NULL	Cooking/cleaning	1	0	0
71	71	1	Walking	0	0	0
81	-1	NULL	Refused	1	-1	-1
82	-2	NULL	Don't know/Can't remember	1	-1	-1

Table 7 A few rows from the activityassetstbl with relevant columns

4.2.4.5 User Data

The user related data for managing the use of the instrument by the users are defined as the user data. This includes the interviewer information, respondent information, interview information, and associated session information. Currently since the system is working in the interviewer-assisted mode, the interview information has a relationship to both the interviewer and the respondent – when the system moves towards including self-administered mode, the relationships to both the interviewers and the respondents will be switched to being independent thus easily enabling multi-mode working. The tables associated with the user data is shown in Figure 18.

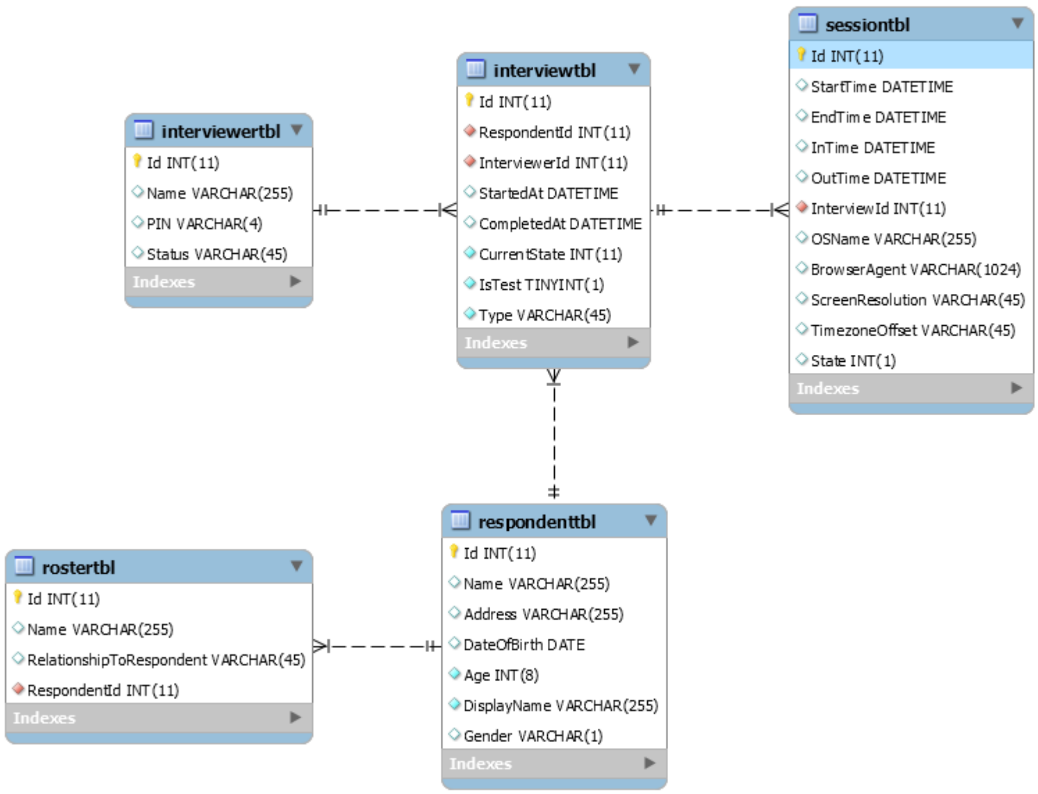

Figure 18 Tables associated with the user data

Currently the rostertbl is not in use, but serves as a placeholder when integration of the modes is to be implemented and the respondent’s roster information can also be obtained.

4.2.4.6 Experiment Data

The data related to the feedback from the interviewers after using the instrument on completing an interview is referred to as the Experiment data. The interviewers are presented with a simple questionnaire based survey after completing every interview where they are asked to indicate their satisfaction with using the instrument. The full questionnaire is available in the Appendix 7.3 The associated tables for storing the experiment data is shown in Figure 19.

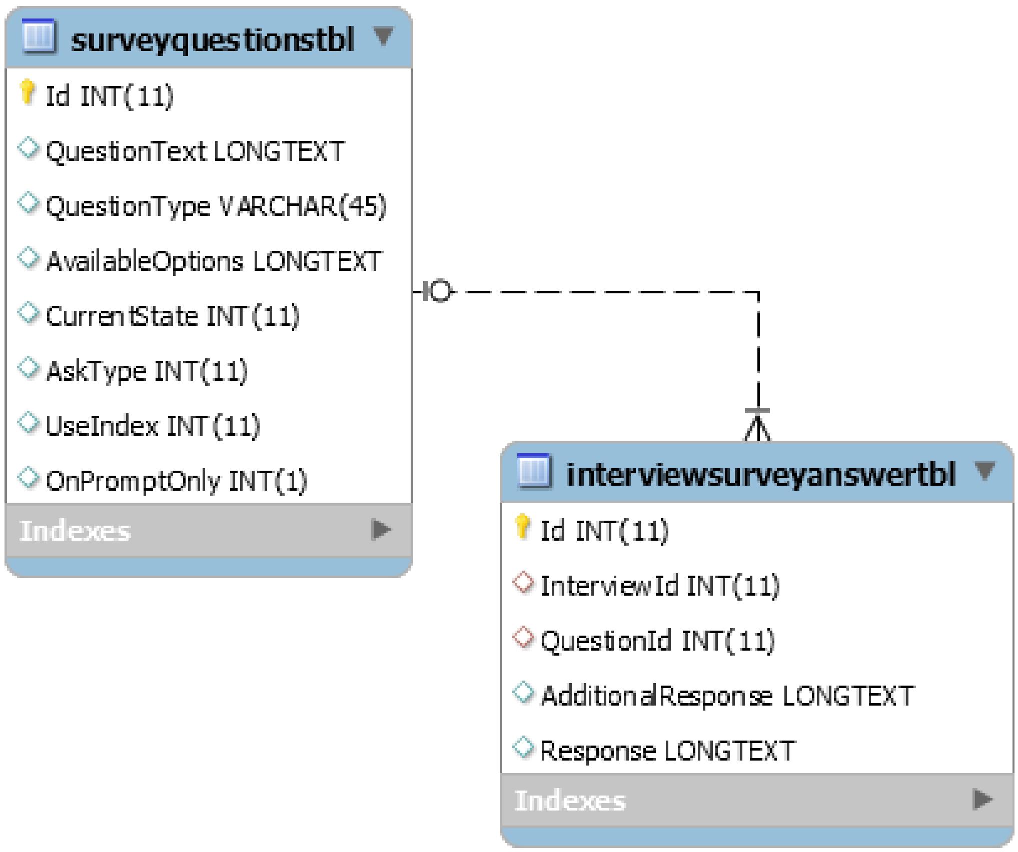

Figure 19 Tables associated with storing the experiment data

It must be noted here that during Phase 01, there was a technical issue with the server computer due to unknown conflicts between Apache and Glassfish (the web application servers) leading to a failure when the frontend attempted to communicate the interviewers’ survey responses to the backend. This inevitably led to the interviewers’ survey responses not being stored or available for analysis.

4.2.5 Interface design

The Graphical User Interface (GUI) for the web application is written in HyperText Markup Language (HTML), with Cascading Style Sheets 4 (CSS4) for the design and JavaScript for the working code. The design was done keeping in mind that there would be mechanisms introduced later on and that the existing mechanisms could undergo changes based on analysis and feedback of the instrument usage. Experts in time diary design were consulted to vet the usefulness and usability of the design. Dr. Robert Belli (Psychology, UNL) and Dr. Don A. Dillman (Department of Sociology and The Social & Economic Sciences Research Center, Washington State University) both independently approved the design with positive feedback. Changes that were required to bring about consistency were also taken and incorporated into the final design of Phase 01.

The ATUS web instrument prototype consists of multiple screens that would guide the interviewer (presently) and later on the respondents to the instrument screen. The instrument screen here refers to the actual page that the user would use to provide their responses also known as the instrument. While the supplementing pages are not directly part of the instrument, they are part of the web instrument as a whole and serves to provide the required resources to prepare the instrument.

Figure 20 shows the flow chart of the various pages and how they are connected.

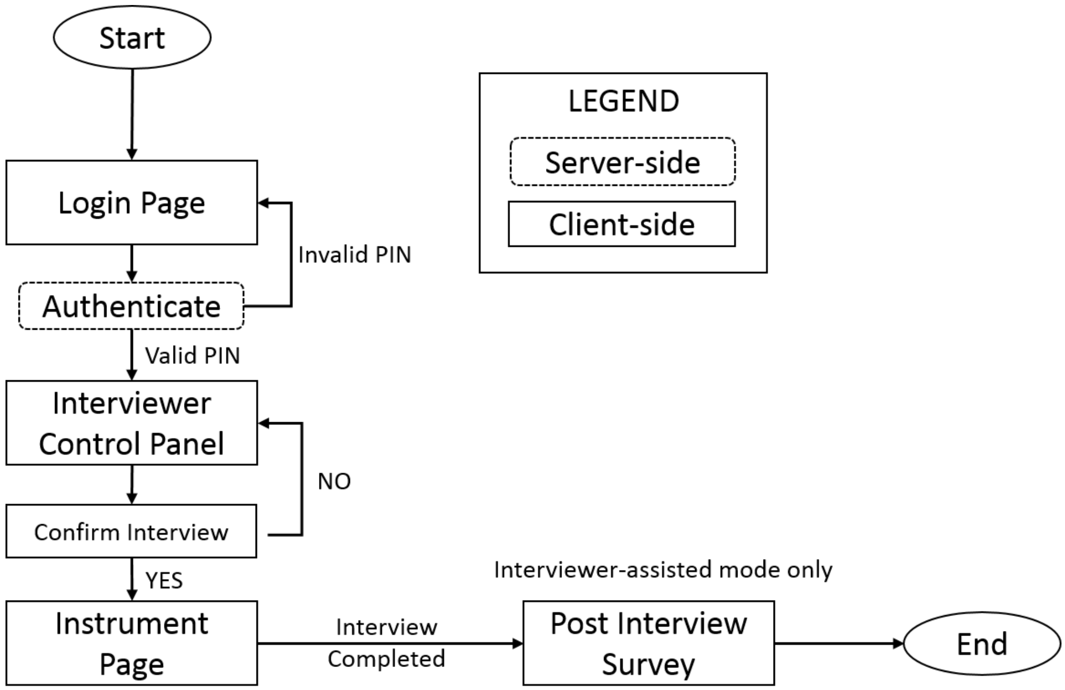

Figure 20 Flowchart illustrating the interviewer's overall actions while using the system

4.2.5.1 Login Page

The landing page or the home page for the ATUS web instrument is the Login page. Here the user would provide their Personal Identification Number (PIN) that was assigned to them ahead of time. Currently, the authentication uses a simple PIN based approach, but this can be easily extended to a stricter username and password based approach over secure channels. The Login page is shown in Figure 21. The user can enter their 4-digit unique PIN in the field and press the Go button to authenticate into the system. Each of the interviewers were assigned a different PIN and the system uses the PIN entered to identify the interviewer logging into the system.

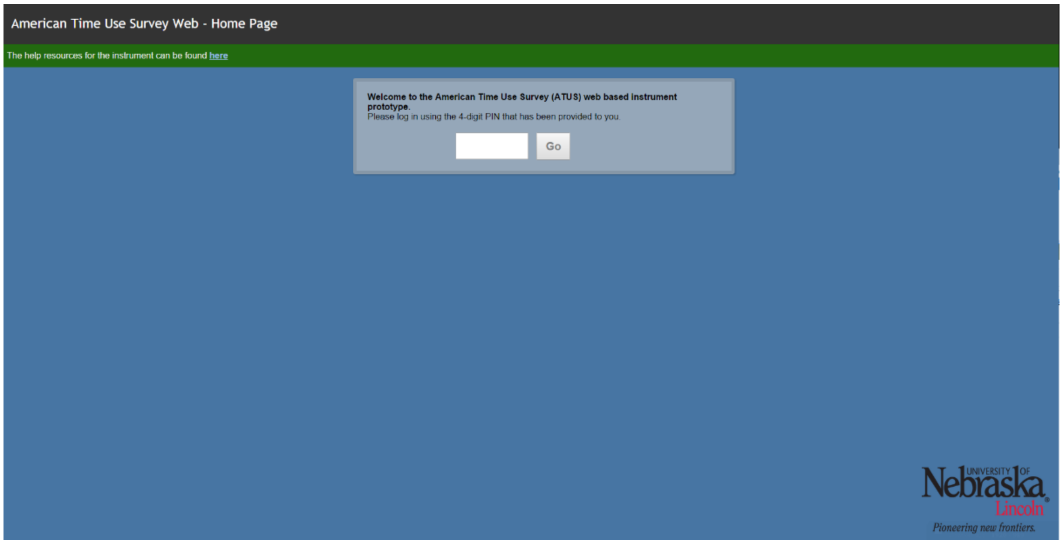

Figure 21 Screenshot of the Login page

The Login page (and the subsequent non-instrument pages) provide a direct link to the resources available to the user to learn how to use the instrument.

4.2.5.2 **Resources Page**

As mentioned, a direct link to the resource page is available to the user in almost all the non-instrument pages. The Resources page provides access to the ATUS user manual, the ATUS interviewer manual and five videos that walkthrough using the different features of the instrument. The videos currently list the following:

1. General Overview Video

    This provides a general use-case scenario of the instrument explaining how the interviewer would login and access their interviews. It also provides a walkthrough of how to enter the activities, provide the context information and edit and delete the activities.

2. Timeline Usage Video

This video shows how the interviewer can use the timeline control to help visualize the activities recorded and some general actions that can be performed such as changing the activity duration, changing the activity start time and selecting and deleting the activity from the timeline itself.

3. Predictions Usage Video

    This video provides a walkthrough on how to use the predictions made by the system (if the interview is a PROMPT type).

4. Overlap Handling Video

    This video explains how the system provides the user with a warning when an activity is entered that overlaps another recorded activity (or activities) time.

5. Missing Travel Prompt Handling Video

    This video shows the warning dialog issued to the user when the system detects a change in the location between two consecutive activities without an identifiable traveling activity between them.

The resource page screenshot is shown in Figure 22.

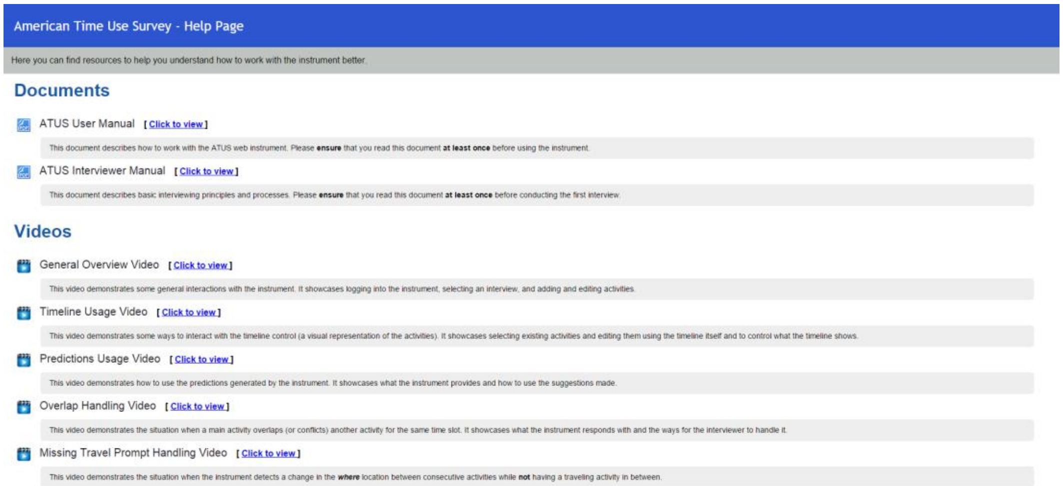

Figure 22 Screenshot of the Resources Page

4.2.5.3 **Interviewer Control Panel**

When the interviewer successfully logs in using their PIN, they are taken to the Interviewer Control Panel, where their active interviews (if any are available) are listed in a tiled manner. The Interviewer Control Panel screenshot is shown in Figure 23.

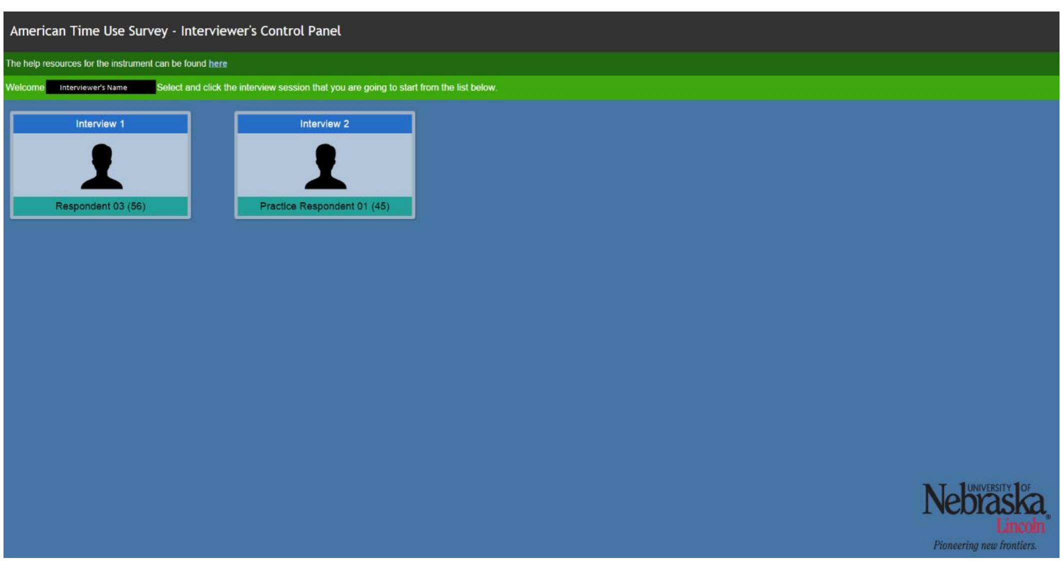

Figure 23 Screenshot of Interviewer Control Panel

The interviewer thus has an easy overview of how many interviews they have pending. When the integrated framework is implemented for SAM, a similar page for the respondent would provide them with their sessions directly where they can record their responses. The interviews being displayed as tiles provides the interviewer with some information regarding the respondent such as their gender and age so as to allow themselves to prepare for the interview. As the system scales, when an interviewer may have many interviews, a simple search box can help them easily narrow down their required interview tile.

To begin an interview, the interviewer can click on the tile representing the respondent for that interview. The system prompts a confirmation dialog that the selected interview is about to start; which when the interviewer confirms, would start the interview by taking the interviewer to the Instrument Page.

4.2.5.4 **Instrument Page**

The Instrument page, as the name suggests, refers to the web ATUS instrument. This page allows for the interactions between the user (interviewer in the current implementation – later on both the interviewer and the respondent) and the system. The instrument page is shown in Figure 24. The instrument GUI is divided into four panels:

1. The instructions panel
2. The input panel
3. The status panel
4. The timeline panel

Figure 24 is the screenshot of the instrument as it is for Phase 02. For Phase 01, the instrument also had a prompt panel (refer to Appendix 7.4) which was removed for Phase 02 after observing that the interviewers did not use it. Each of these panels generally contains an implementation of an Interaction Mechanism in them. For example, panel 1 contains the Precode Interaction Mechanism, while panel 2 contains the Autocomplete Interaction Mechanism.

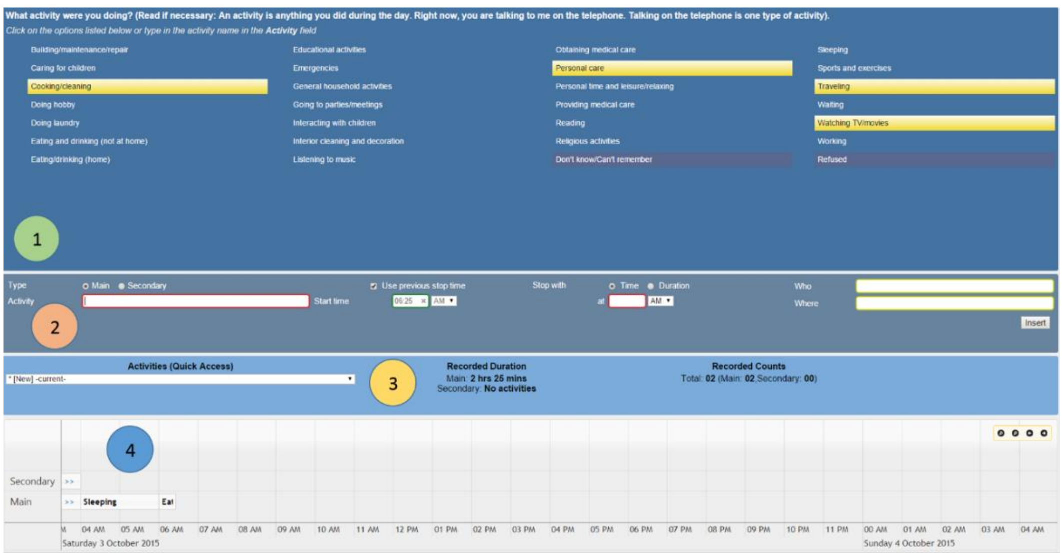

Figure 24 Marked screenshot of the Instrument page during use

If the interview is a PROMPT type interview (as is in Figure 24), the predicted precodes in the instruction panel would be highlighted as shown. In a NO-PROMPT interview, this highlighting would be absent.

4.2.5.4.1 Panels

The instrument page consists of different panels (separated visual sectioning) that build up the instrument as a whole.

4.2.5.4.1.1 Instructions Panel

The instructions panel displays the instructions and the precodes to the interviewer in a listed format. The generation of the precode was part of the design time KE and incorporates domain knowledge into the system. The content of this panel changes to suit the current action being performed by the user. These include the current field being edited, the state of the interview, and hard and soft warnings. The Precode Interaction Mechanism is implemented via this panel, wherein the precode list is displayed and controlled through its client side implementation.

The precode list in the instruction panel lists the precodes – the set of items that is identified as being important or common – that the user can click on for entry into the corresponding field. This click and enter approach is intended to make entry of routine information faster without having to memorize other indexing methods (such as a precode number in ATUS). The instruction panel individually is shown in Figure 25. The screenshot is taken from a PROMPT interview and thus the predicted next activity is highlighted. Figure 25 is also the working implementation of the predictions being rendered by the Precode Mechanism (as opposed to a separate Prediction Panel in Phase 01).

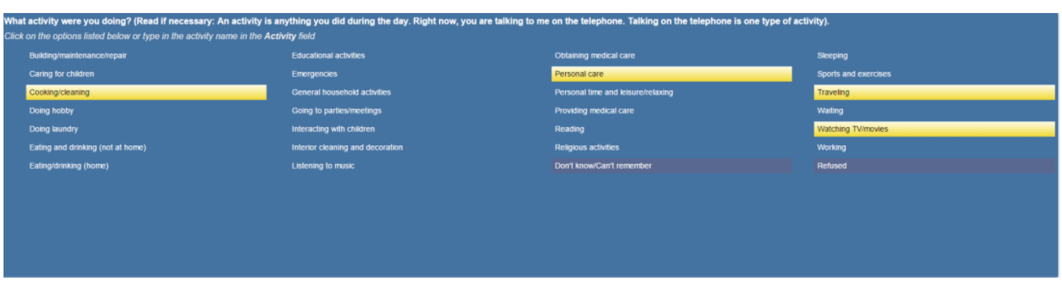

Figure 25 Zoomed in view of the Instructions panel

4.2.5.4.1.2 Input Panel

The input panel (Figure 26) displays the fields in which the information regarding an activity can be entered. This includes the type of activity (Main or Secondary), the Activity name, the Start Time of the activity, the Stop time type of the activity (time or duration), the fields for entering the stop time or the duration in hour and minutes, the context information fields for Who and Where and an Insert/Save button. If an activity is being edited, a Delete button appears next to the Save button. The Autocomplete Interaction Mechanism is implemented on the three text input fields (Activity name, Who, and Where). The content for the autocomplete is loaded when the instrument initializes.

Figure 26 The Input Panel of the instrument zoomed in

4.2.5.4.1.3 Status Panel

The status panel (Figure 27) displays information regarding the current status of the interview and a list view of the activities entered called the Quick Access dropdown. This allows for easy selection of those activities that are usually too small to be seen and selected from the timeline. The status panel shows the total recorded durations of the main and secondary activities and their counts. When an interview is completed (the 24-hour period of main activities is recorded), a Finish Interview button appears on the status panel for the interviewer to confirm and complete the interview.

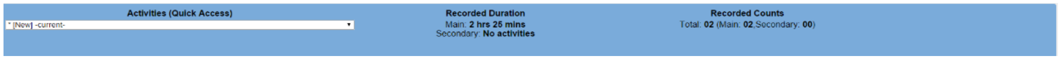

Figure 27 Zoomed in view of the Status panel

4.2.5.4.1.4 Timeline Panel

The timeline panel contains the timeline control (Figure 28) and is the bottom-most panel. It provides a chronologically arranged view of the activities that have been recorded and also allows some editing of the activities such as changing the start time, stop time, duration and deleting the activity. The timeline serves as one of the main locations from which the user can select an entered activity for editing at a later point of time. The control supports 4 view actions – zoom in, zoom out, scroll left and scroll right. These actions can be performed either by using the corresponding buttons at the top right of the timeline control or by using the mouse wheel and drag. When an activity is selected (either using the timeline or the quick access), they are actively selected on both the controls thus maintaining reliability. This is the implementation of the Timeline Interaction Mechanism.

Figure 28 Zoomed in view of the timeline control

4.2.5.4.2 Dialogs

When there is a warning or error that requires immediate attention from the user, the instrument uses blocking dialogs (where the instrument’s UI is blocked by an overlying dialog box) to present the information to the user. The user must provide feedback to the dialog before continuing with normal usage. Currently, the instrument raises dialogs for two purposes:

1. Missing travel
2. Activity overlap

4.2.5.4.2.1 Missing travel dialog

When the instrument detects two time adjacent activities being recorded with different locations without a traveling activity between them, it raises the Missing travel dialog informing the user (the interviewer) of this detection. The user can then resolve it by requesting the information from the respondent (in interviewer-assisted mode) or attempt to recollect and enter the information (in self-administered mode). When self-administered mode is implemented though, this dialog will undergo suitable changes to make it more informative and assistive to the respondent based on the mechanisms changes. An example missing travel dialog being raised is illustrated in Figure 29.

Figure 29 Zoomed in (and cropped) view of a Missing travel dialog

The missing travel dialog informs the user that there is a difference in the 'Where' location of two activities and then displays the two activities with their times for clarity.

4.2.5.4.2.2 Activity overlap dialog

When the user attempts to add a main activity at a time duration that already has one or more main activities recorded, the instrument raises the Activity overlap dialog to bring attention to this. Furthermore, the Activity overlap dialog will also provide with a resolution method of splitting the attempted activity to fit into any gaps (if available) and then keeping the overlapping parts as secondary activities. The user has the option to either accept this resolution method or to attempt to fix it on their own. An example of the Activity overlap dialog is shown in Figure 30. Thus, this is the implementation of the visual display of the Overlap Handling Mechanism.

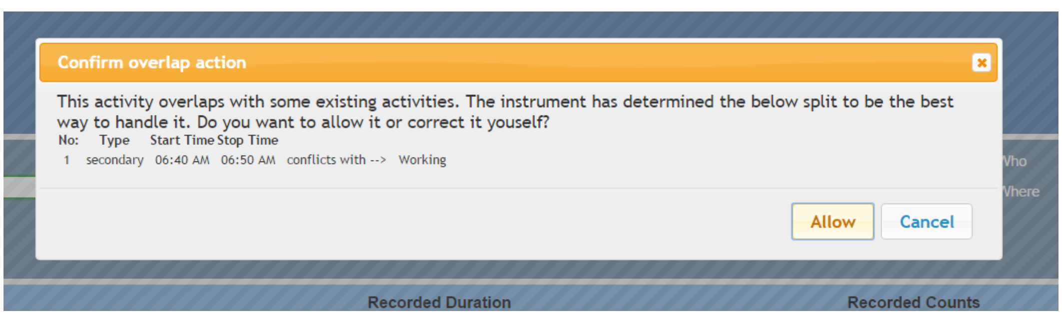

Figure 30 An example of the Activity overlap dialog

4.2.5.4.3 Warnings

The instrument also employs non-intrusive warnings to bring the user's attention to missing or invalid information in the data currently being entered. Warnings are raised only when the user attempts to add or save an activity. The instrument uses two types of warnings:

1. Hard warnings – These warnings inform the user that there is invalid or missing information that is required before the add or save action can be performed. The user cannot ignore this warning and must suitably address them by correcting the data before proceeding. The validation errors are shown to the user as hard warnings in this implementation as illustrated in Figure 31.

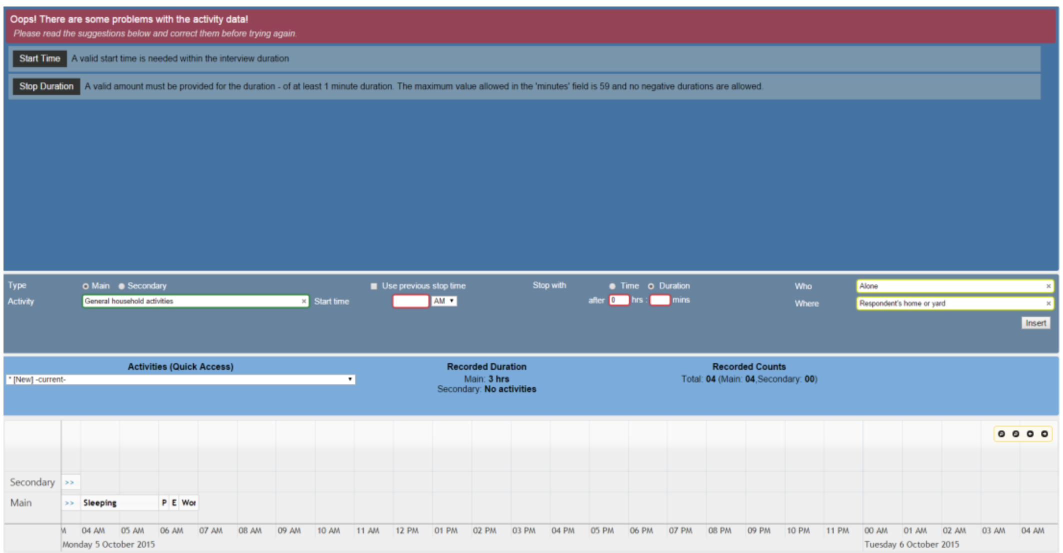

Figure 31 Hard warning being raised due to missing start time and duration

2. Soft warnings – These warnings inform the user that there is some information missing in the current activity when the add or save action is being attempted. The user can choose to ignore this warning and proceed with information missing or resolve it using the options provided. In the current implementation, missing ‘Who’ and ‘Where’ fields are displayed as soft warnings for activities that optionally require it as shown in Figure 32.

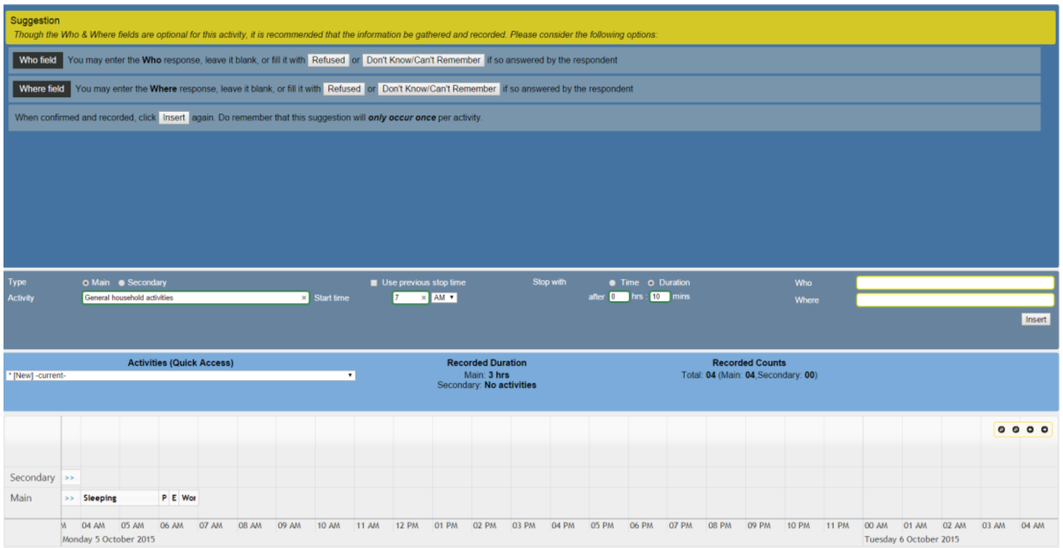

Figure 32 Soft warning being raised due to missing 'Who' and 'Where' for an activity where it is optional

4.2.5.5 Post Interview Survey Page

When the interviewer completes the 24-hour period of the time diary and confirms the completion of it in the instrument page, they are redirected to a Post Interview Survey page, where the interviewer is requested to answer some questions regarding their use of the instrument. The full list of the questions asked to the interviewer is listed in Appendix 7.3. A screenshot of the page is shown in Figure 33. This survey is intended only for the interviewers to gauge their feedback about the instrument.

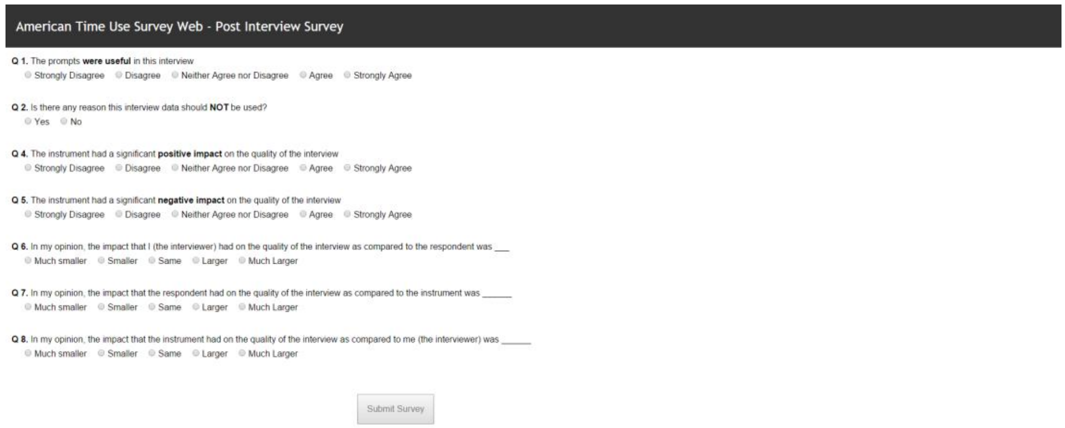

Figure 33 Screenshot of the Post Interview Survey page

4.2.6 Implementation

As the system is implemented in the client-server model with a rich client side implementation, the server side implementation focuses on handling the data processes and the intelligent processes of the system. The data processes involve recording the data received from the client-side, retrieving and forwarding the data requested by the client-side while the intelligent processes involve the actions relating to the Knowledge Engineering mechanisms such as making the predictions. As more Knowledge Engineering or Interaction mechanisms are implemented, the server-side is where the part of the online implementations that need access to data would reside while the parts that use the data to interact with the user would be on the client-side.

4.2.6.1 **Server Objects design**

The server-side currently contains the implementation of the data recording components and the implementation of the prediction mechanism. The server-side part of the prediction mechanism uses the knowledge-engineered data and the response data (for context) to generate a set of predictions for the next activity. The server-side consists of the following object structure:

1. AgentBase

   The AgentBase is the base class that defines an Agent for the system. Further implementations of the Agent would thus inherit from this Class. The primary parts of the AgentBase are defined for managing multiple agents using a central controller and for the various endpoint methods for the client-side to access which would trigger the various mechanisms. Currently, implementations of the prediction mechanism and the missing travel mechanism are accessible through the AgentBase.

2. AgentCommunicator

   This class acts as the server-side endpoint for the calls from the client-side wherein it would redirect the call to the appropriate agent handling this interview instance. The data (if any) that is generated by the call would then also be rerouted to the client-side using this class.

3. Recorders & Managers

The Recorders refer to the classes that are involved in handling the data recording calls from the client-side. This involves the response data being generated and recorded by the user and the paradata collected while the user interacts with the instrument. Managers refer to the classes that act as duplex communication channels for the data that is required by the client-side to run properly. This mainly involves the Auxiliary data and the user data.

4. Misc. Classes

   The miscellaneous classes involved in the smooth running of the server-side includes the Entity classes (to access the database), the com classes (to send and receive data from the client-side) and the system management classes that create and manage agent instances for the interview sessions.

4.2.6.2 **Workflows & Mechanisms**

The operation of the instrument attempts to provide a structured flow for data entry with allowances for multiple data entry methods, where the Interaction Mechanisms work to provide data to the interviewer and gather data from the interviewer when they interact with the mechanisms. These interactions are used to invoke the appropriate Knowledge Engineering Mechanisms. To ease the cognitive load on the interviewer with respect to the amount of data they have access to, most of the fields within the instrument can be filled out in multiple ways. Different Interaction Mechanisms interact with the interviewer differently. The following section compounds on these workflows when the interviewer uses the instrument and draws together the operation of the instrument as a whole by describing a flow of the work done.

Referring to Figure 34, the following workflows are the internal flowcharts within the 'Instrument Page' block. At this point, the interviewer has logged into the instrument's main page and is in the process of getting in communication or already in communication with the respondent via telephone. During the use of the instrument for the interview process, a workflow represented by Figure 34 is in place. It shows the actions within the system by the interviewer (seen as the interviewer's interaction), and those performed by the system and the mechanisms.

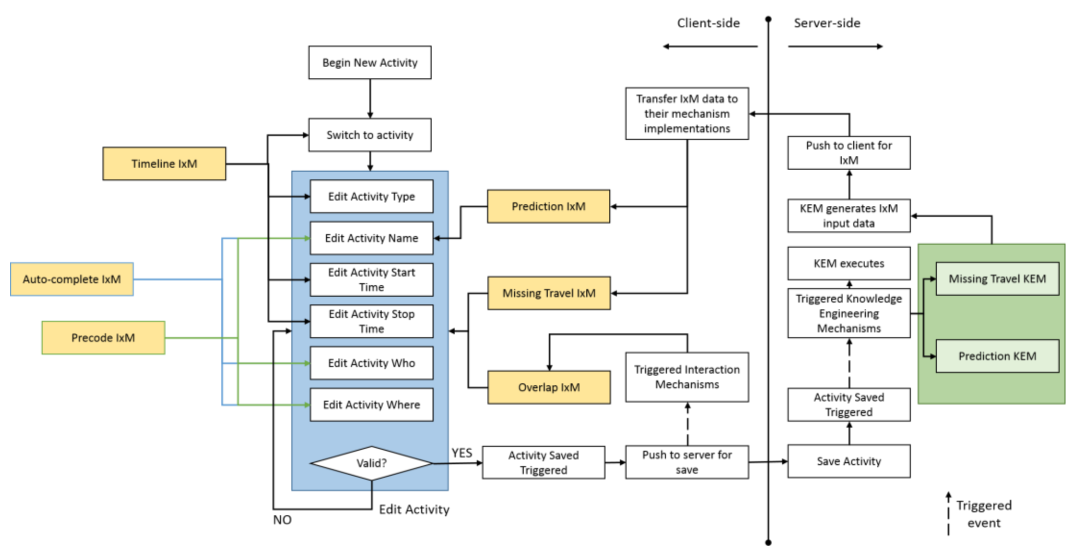

Figure 34 Internal workflow during instrument use

The first new activity is created when the interview starts, after which a new activity is created every time an activity is created and added. For every new activity, the interviewer extracts the required information from the respondent via conversation and inputs the data into its corresponding field. The information required for an activity are:

1. Activity Type – This can be main or secondary. Mandatory.
2. Activity Name – The name of the activity. If the interviewer can determine a coded activity that corresponds to the verbatim response provided by the respondent, they may enter that otherwise they can choose to enter the verbatim response itself. Mandatory.
3. Activity Start Time – The time the respondent reported as having started this activity. Mandatory.
4. Activity Stop Time – This represents the time the activity ended and can be provided as a time value itself (Stop Time) or as a duration (in hours and minutes). Mandatory.
5. Who – This is 'who' the respondent performed this activity with. Depends on the activity. Can be mandatory, optional or not-required.
6. Where – This is 'where' the respondent performed the activity. Depends on the activity. Can be mandatory, optional or not-required.

Once all the required information is entered in, the interviewer can proceed to save the activity. At this point, a validation process runs to determine if all mandatory information has been provided and any data that can be validated for type (numbers for duration, valid hours) is correct. If the validation fails on mandatorily required data, hard warnings are shown to the interviewer indicating the missing data. If the verification fails on non-mandatory data, the corresponding soft-warnings and dialogs are displayed to the interviewer from where the interviewer can decide on how to proceed. Once the duration of the reported and recorded activities satisfies the required 24-hour duration during the previous day, the interviewer is provided with a button to confirm the end of the interview's data recording process which when clicked completes the 'Instrument Page' block in **Error! Reference source not found.** and proceeds to the next page (Post-I nterview Survey for the interviewer).

During the course of the data entry process, the events raised by the interviewer interacting with the instrument is used to drive the system's actions. An event is defined as a particular condition or state being reached. Thus for example, 'Activity Saved' event would occur when an activity is saved. By using events on the client-side and the server-side, the corresponding KEMs and IxMs are executed to perform their actions. In this implementation, the 'Activity Saved' event is used to start the Prediction KEM to generate a list of predictions at real time based on the previous activity entered by the interviewer.

Once the interviewer saves an activity and it has been validated, it is sent to the server to be saved. This raises the 'Activity Saved' event on the server side, as mentioned before, the Prediction KEM is executed. This results in the creation of a list of possible next activities – Top 5 in each method for Phase 01, Top 5 using Previous Activity Based method in Phase 02, which is then sent over to the client side where this data is passed to the Prediction IxM. The Prediction IxM then displays this list to the interviewer ordered by the probability (Phase 01). For Phase 02, since the Prediction IxM shifted to using the Precode Mechanism to deliver the predictions, the ordering of the Precodes' determine the display order of the predictions. Each of the mechanisms in the system attempt to provide the interviewer with an alternate means of data entry to reduce the data entry time and/or also attempts to reduce the cognitive load by presenting information in a more concise and clear manner.

The data used by the mechanisms includes Auxiliary data and the Response data. As part of determining the Auxiliary data for the system, Knowledge Engineering was performed while designing the system – this was termed Design Time Knowledge Engineering and is described in detail in Section 4.3.

4.2.6.2.1 Precode Interaction Mechanism

The precodes are a predefined list of items that are deemed most likely to be reported by the general population. Defined originally in the ATUS code book, it served as a quick list for the interviewer to refer to when the respondent reports the activity, who and where information. In our implementation, we borrow the idea of having the precodes as part of the Precode Interaction Mechanism. As part of the design time KE, this list was modified to fit the other Auxiliary data of the system (to include the activity mapping etc.). The precode interaction in the current implementation is intended to provide a click-to-use ability for the interviewer wherein, the interviewer can click on a precode option to fill it into the corresponding data field. This is expected to reduce the time taken by the interviewer to fill in the data, thus decreasing the overall interview time and could also help the interviewer keep the conversation going with the respondent. The full list of all the precode options are listed in Appendix 7.5. The interviewer's interactions with the precode IxM are tracked and recorded and can be used as the starting data for future implementations where it can serve as training data for determining how it is used by the different types of users.

4.2.6.2.2 Timeline Interaction Mechanism

The timeline control of the instrument is part of the Timeline Interaction Mechanism. It serves to display the respondent's reported activities of the day in a chronological order with basic editing options. This is intended to help the interviewer visualize the activities reported easily to assist the respondent with their recollection process and to gain an idea of how far long they are. Presence of progress indicators in surveys have been linked to increased response rates but have not yet been proven to have significant effects directly (Couper, Traugott and Laminas, 2001). In our methodology, we view the timeline as a visceral progress indicator to allow interviewers get a quick glance about the data recorded so far. The paradata for the timeline interactions are also tracked and recorded for future use.

The timeline also provides with some basic duration-based edit features and the ability to switch between activities quickly. The interviewer can select any recorded activity at any point of time during the interview to load up that activity's details quickly. The timeline also provides a simple click-and-drag feature to change the start and end time of an activity. It also provides with a simple way to delete a selected activity and to change the activity between primary and secondary using a drag and drop feature. The intention of the Timeline Interaction Mechanism is to reduce the cognitive load on the interviewer and to decrease the interview time by providing assistance to the interviewer for using the instrument.

4.2.6.2.3 Autocomplete Interaction Mechanism

The Autocomplete Interaction Mechanism introduces an autocomplete feature into the data entry process of the instrument. The autocomplete feature is activated when the interviewer begins typing in a data field and it brings up a list of alphabetically ordered items that contain the characters entered by the interviewer. If the interviewer types in more characters, the filter is extended to include those characters effectively narrowing the options down. While the autocomplete feature is a standard in many web applications, its use during IAM helps the interviewer perform two actions at once – search through the auxiliary data for a coded activity, who or where item and to attempt to convert the respondent’s verbatim to a coded item on the fly. This information is also tracked and recorded and it can be extended for future use where the filters can also include predictive and suggestive items further helping the interviewer narrow down the coded response for a verbatim response and reduce the data entry time.

4.2.6.2.4 Missing Travel Mechanisms

The Missing Travel Mechanisms is a dual-part mechanism. A dual-part mechanism has two parts to the overall mechanism – usually a KEM part and an IxM part. Each of the parts are individually referred to as the Mechanism’s KEM and IxM part. The dual-part is needed to distinguish the component separation that generates the data required for its counterpart. The Missing Travel KEM part is executed by the Activity Saved event. It executes the check for missing travel information as shown in Figure 35.

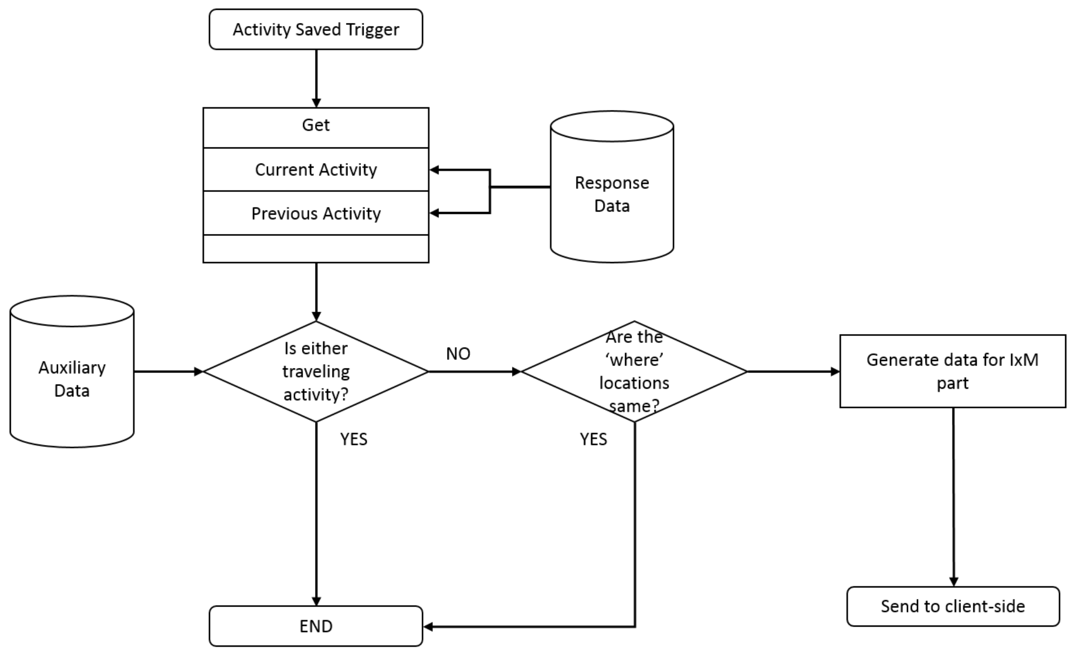

Figure 35 Missing Travel Mechanism flowchart

Once the missing travel data is generated on the server-side it is sent to the client-side to be used by the Missing Travel IxM part to issue a dialog to the interviewer that there is a missing travel between the two activities. The interviewer can then further probe the respondent regarding the missed out activity and record this. This is thus intended to increase the data quality of the responses obtained using our instrument. For future extensions, it can be used to notify and assist the user (interviewer and/or respondent) about the missing travel and allow them to provide the response in a faster click-to-enter form.

4.2.6.2.5 Prediction Mechanisms

Similar to the Missing Travel Mechanisms, the Prediction Mechanisms is also a dual-part mechanism. It consists of the Prediction KEM on the server-side and the Prediction IxM on the client-side. The Prediction KEM is executed by the Activity Saved event on the server-side and initiates the process of generating a list of activity predictions for the next activity. There were two methods of this prediction in Phase 01 – the Previous Activity Based (PAB) method which predicted the activity based on the activity just before it and the Time of Day (TOD) method which predicted the activity based on the time of the day the activity is starting at. The TOD method also considers the day of the week in its prediction determination. The process flowchart is shown in Figure 36.

When the Prediction IxM receives the predicted next activity data, it displays this list to the interviewer either as a prompt panel (Phase 01) or using precode highlighting (where the corresponding precode entry is highlighted) (Phase 02). The interviewer can then click on the predicted option and it is filled in the corresponding data entry field. This is thus intended to provide the interviewer with an easy recommendation list of sorts though within our system we call these the predictions since recommendations would imply a suggestive relationship which is not allowed in surveys.

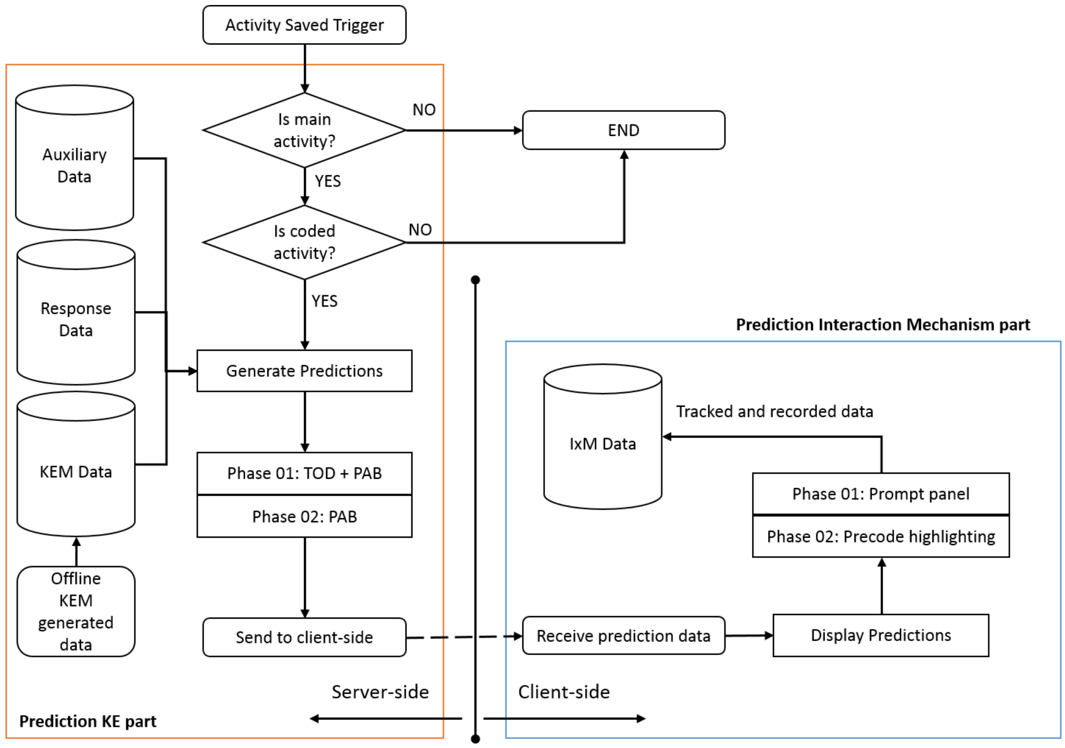

Figure 36 Prediction Mechanisms flowchart

By providing such predictions, we hope to help the interviewer reduce their data entry times and passively increase the data quality by allowing the interviewer to maintain their conversation with the respondent with the least interference by data entry. The two prediction methods use the data from the KEM data (described in Section 4.2.4). This data was generated using KE processes done offline using the ATUS data and is described in detail in Section 4.3. Moving this generation to an offline KEM helps reduce the computation time for the prediction by very significant amounts since it is converted to a lookup rather than a search.

4.2.6.2.6 Overlap Handling Interaction Mechanism

The Overlap Handling Interaction Mechanism is a client-side IxM that provides assistance to the interviewer when the system detects an attempt to save an activity that overlaps with another already recorded activity. By making this an IxM rather than a simple validation process, we provide the interviewer with easy ways to resolve the overlap rather than forcing the interviewer to always detect the overlapped activities and apply themselves to fix the overlap. This is thus intended to reduce the cognitive load on the interviewer during an overlap situation. The current implementation of the Overlap Handling IxM describes the overlap situation and provides a resolution option where the current activity is split into multiple smaller activities that are made secondary over the existing overlapped activities as shown in Figure 37.

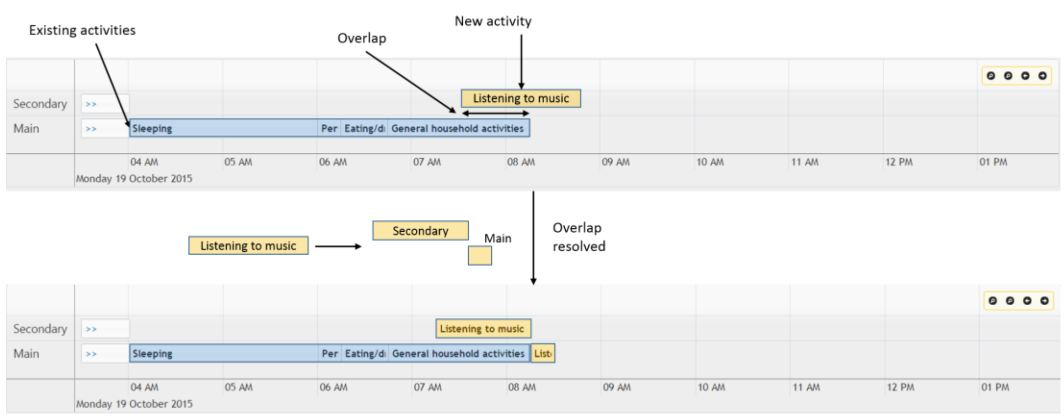

Figure 37 Overlap handling illustrated

A new activity overlaps an existing activity (or activities) if their start time and/or end time cause them to be occurring within a shared time frame. To resolve this, the system splits the new activity into as many parts as there are shared time frames, with each part having a start time at the start of the shared time frame and end time at the end of the shared time frame. These parts are then changed to be of secondary activity type. For future extensions, the overlap handling mechanism can be extended to include more resolution options and a paired KE mechanism that recommends the best resolution method based on the context. This can be extended for both SAM and IAM. It must be noted here that currently two main activities are not allowed to cover the same duration, but more than one secondary activity can. If the respondent reports more than one activity for the same duration, the interviewer may ask the respondent to pick the one that they think is the main activity.

4.2.7 Data Generation

The domain information for time diary surveys consists of the coded activities, who and where context information that can be recognized by the system, the instructions and messages to be shown to the user and the associations between the context information and the activities (which activity needs what context information). To enable the system to be extensible, all of the Knowledge engineered must relate to the domain information and thus the domain information created is called the Auxiliary data. Since the instrument is inspired by the ATUS instrument, we used the coded activities provided by the ATUS codebook as the starting point. This consists of 347 activities at the finest level (refer to appendix 7.1 for the full listing). The original coding by ATUS defines three tiers of activities called Tier 1 (T1), Tier 2 (T2) and Tier 3 (T3) in increasing granularity. Thus higher (3 is the highest, 1 is the lowest) tiers are grouped together to create the lower tier level. This grouping is illustrated in Figure 38. There are 18 Tier 1 codes, 110 Tier 2 codes and 347 Tier 3 codes (468 in total including codes for non-coded activities). While this listing provided us with an almost comprehensive list of activities, the wording and the nature of the grouping was deemed too precise to be usable in a conversational setup. Certain activities like Exterior repair, improvements, & decoration (code 020402) would require the user to mentally convert their activity into the coded form and with the amount of granularity introduced by the 347 activities, would require some time to be narrowed down. Indeed, in the ATUS process, the interviewers simply record the verbatim response during the interview and a lengthy coding process by trained coder personnel converts the verbatim response into their corresponding Tier 3 activity.

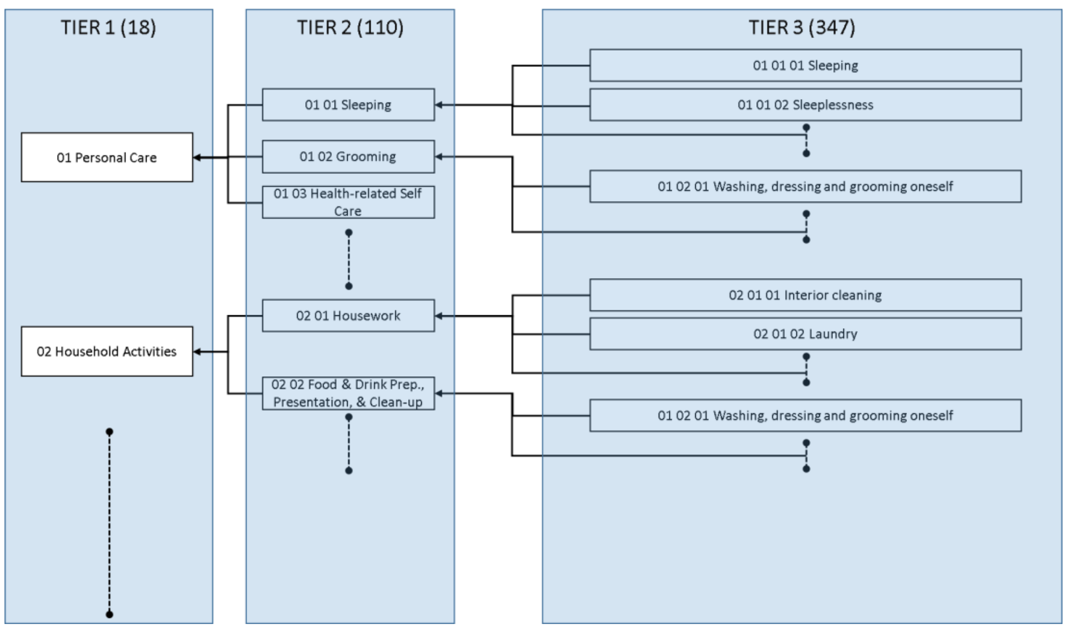

Figure 38 Tiered structure of the ATUS' coded activities

Thus understanding the unsuitability to use the coded activities directly from ATUS, we devised an alternate set of narrowed down activities and introduced a tiered grouping similar to ATUS. For this, the 347 T3 activities from ATUS were taken and reduced to a set of 80 activities called the MID tier. We also introduced the idea of a Concept which is the name for a set of both MID and T3 activities that fall under a general umbrella. Both the MID and T3 were mapped to the Concepts so as to compare the effectiveness of using our grouping over the one by ATUS and to introduce the ability to modify and extend these at later stages. The Concept mapping was then denoted by L-Concept and D-Concept based on how they relate to the T3 activities. Here L-Concept is a shortening for Linked-Concept and D-Concept for Direct-Concept. The grouping is illustrated in Figure 39. While cursorily it would seem redundant to have an apparent replication of the three tiers for activities from ATUS, our tiers perform operational roles, as opposed to the coding role for the ATUS tiers. The concepts are used to handle the sparsity in the T3 activities and to bring related activities under one 'concept' from which predictions can be made. The introduction of our tiers (which is based off ATUS's T3 activities) allows us to cover the same range of activities as defined by ATUS while also providing us with the flexibility to name them in a way that would make more sense for the user to understand.

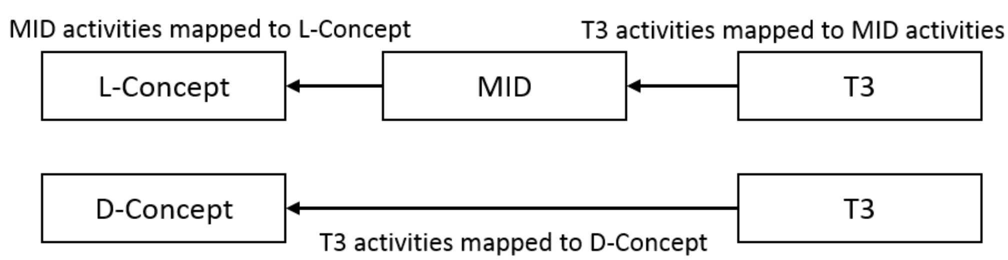

Figure 39 Tier 3 (T3) activities and their mapping to MID, L-CONCEPT and D-CONCEPT

The thus defined MID activities were then used as the coded activities for the implementation. The MID activities were worded to be more encompassing and simple and the reduced number of them allows for easier narrowing down of the verbatim responses. The resulting transformation and the process was approved and vetted by Dr. Robert Belli, an expert in time diary surveys.

4.3 Design-time Knowledge Engineering

During the design of the mechanisms, as part of the instrument design on the whole, Knowledge Engineering was performed on the ATUS response data to determine the usability of the activity predictions created from it and the effectiveness of different Machine Learning techniques that could be used to predict the next activity given the previous activity. This Knowledge Engineering process came to be known as *Design-time Knowledge Engineering* (DTKE) since it was primarily done offline and verification was performed manually under the guidance of experts such as Dr. Belli. The resulting information from the Design-time Knowledge Engineering was hugely influential in determining the implementation of the data storage and the data models used.

The DTKE was primarily focused on determining how to predict the activity that the respondent was about to report. This is the offline KE mentioned earlier that generated the prediction tables that would be used by the Prediction Mechanisms. For the prediction, as mentioned previously in Section 4.2.6.2.5, two methods were devised:

1. Previous Activity Based (PAB)

   This prediction method would use the activity that was previously entered (in chronological order) to predict the next activity that could be reported. For this method, three machine learning techniques were studied, trained and tested on the ATUS historical database (for the years 2010, 2011, 2012, and 2013). The total number of activities recorded in successful interviews on record are given in Table 8.

Year	Total Activities
2010	257,193
2011	234,358
2012	230,920
2013	215,567
Total	938,038

Table 8 Total activities for the ATUS data between 2010 and 2013

2. Time of Day (TOD) based

   The Time of Day (TOD) prediction method used the historical data from ATUS (2010-2013) to generate a probability ordered list of activities that occur at each 30-minute mark during the day. This also takes into account the day of the week. Thus during the offline KE, the activities reported by the respondents on each day of the week (Sunday through Saturday) were taken and the probability of each activity occurring at every 30-minute mark (04:00, 04:30, 05:00, 05:30 etc.) were calculated. From this, the top 5 activities were taken as the predictions for the activity at a given time (adjusted to its 30-minute mark).

4.3.1 Previous Activity Based (PAB) Prediction

To generate the prediction list for the next activity given the previous activity, we investigated machine learning algorithms that could predict sequential items. The machine learning methods that were used are:

1. Markov Chain Models (MCM) (Bishop, 2006)
2. Artificial Neural Networks (ANN) (Yegnanarayana, 2009)

In the process of the KE, we also conducted Principal Component Analysis (PCA) on the demographic attributes of the respondents in an attempt to identify which of the respondent demographics (if any) would affect the predictability. The respondent demographics were obtained as part of the ATUS historical data and consist of a reduced set of 69 attributes such as age, gender, race and ethnicity, income related attributes, etc. These are collected from the respondent a few weeks ahead of the actual time diary survey as part of another survey called the Current Population Survey (CPS) via a telephone interview. Based on the analysis of the results from testing the learning algorithms, we generate the prediction list used by the Prediction Mechanism's Knowledge Engineering Part in Section 4.2.6.2.5.

4.3.1.1 **Training/Testing Methodology**

The ATUS historical data consists of many sets of data such as the response data, CPS data, call history data etc. The response data contains the coded activities reported by the respondent during their time diary survey interview, while the CPS data contains the respondent's demographic attributes recorded during a CPS interview. The historical data from ATUS we consider are from the years 2010 through 2013. Though data exists from before 2010 and now after 2013, we decided to use only data from 2010 to 2013 because the data before 2010 was distinctly different in its structure and coding from the later ones. Since each learning algorithm requires a training and testing data set, we decided to use the data from an entire year as the testing data set and the rest of the 3 years as the training set. We figured that this would be a good way to take the knowledge from one year (training) and test its usefulness against the others (testing). Every coded

Tier 3 activity has a transformation applied to it that converts it into another coded form as mentioned in Section 4.2.7. This was done to overcome the sparseness of the Tier 3 activities' activity-next activity combinations. When the respondent's reported activities are arranged in a chronological order, we recreate the respondent's activity sequence for the day. Each activity followed by its immediate time adjacent activity forms the activity-next activity sequence. To further imbibe the usefulness of the transformation process, we create a set of configurations that completely describes the data used by the machine learning algorithms during testing. A configuration consists of the following parameters:

1. Training data set year
2. Testing data set year
3. A transformation describing the coding format for the activity and the next activity in the activity-next activity sequence.

This creates a total of 60 configurations that were then run through each of the learning algorithms to determine the algorithm that had the best prediction power. The configurations used are listed in Table 9. The full list of all the transformations for the Tier 3 activities are detailed in Appendix 7.6.

Sl. No.	Trained Year	Tested Year	First activity transformation	Next activity transformation
1	2010	2011	D-CONCEPT	MID
2			L-CONCEPT	MID
3			D-CONCEPT	T3
4			MID	MID
5			T3	T3
6	2010	2012	D-CONCEPT	MID
7			L-CONCEPT	MID
8			D-CONCEPT	T3

9			MID	MID
10			T3	T3
11	2010	2013	D-CONCEPT	MID
12			L-CONCEPT	MID
13			D-CONCEPT	T3
14			MID	MID
15			T3	T3
16	2011	2010	D-CONCEPT	MID
17			L-CONCEPT	MID
18			D-CONCEPT	T3
19			MID	MID
20			T3	T3
21	2011	2012	D-CONCEPT	MID
22			L-CONCEPT	MID
23			D-CONCEPT	T3
24			MID	MID
25			T3	T3
26	2011	2013	D-CONCEPT	MID
27			L-CONCEPT	MID
28			D-CONCEPT	T3
29			MID	MID
30			T3	T3
31	2012	2010	D-CONCEPT	MID
32			L-CONCEPT	MID
33			D-CONCEPT	T3
34			MID	MID
35			T3	T3
36	2012	2011	D-CONCEPT	MID
37			L-CONCEPT	MID
38			D-CONCEPT	T3
39			MID	MID
40			T3	T3
41	2012	2013	D-CONCEPT	MID
42			L-CONCEPT	MID
43			D-CONCEPT	T3
44			MID	MID
45			T3	T3
46	2013	2010	D-CONCEPT	MID
47			L-CONCEPT	MID
48			D-CONCEPT	T3
49			MID	MID
50			T3	T3
51	2013	2011	D-CONCEPT	MID

52			L-CONCEPT	MID
53			D-CONCEPT	T3
54			MID	MID
55			T3	T3
56	2013	2012	D-CONCEPT	MID
57			L-CONCEPT	MID
58			D-CONCEPT	T3
59			MID	MID
60			T3	T3

Table 9 The data set configurations used for the learning algorithms

For each of the configurations, the learning algorithms divide the data set (training and testing) into groups defined by each demographic value; for example, all males form a group and all females form a group and each data set is divided into them. These demographic groups are then used as the data set for training and testing respectively. This is thus defined as the *demographic models based testing*. We also perform a training and testing using the entire non-grouped data set and this is defined as the non-demographic model based testing. These two testing types are to further investigate if the respondent demographics have a say in the respondent's activity sequence during the day; intuitively, we hypothesize that the respondent's demographics and the respondent's activity sequence would have a relation. This relationship can be supported if we observe the demographic model based testing significantly perform better than the non-demographic model based testing. Furthermore, if the demographic model can capture the pattern in the respondent's activity sequence of that demographic group, we would be able to observe a significant difference in the accuracy between the demographic model based testing of that demographic against the non-demographic model based testing. This means that the non-demographic model would not perform as well as the corresponding demographic model based testing since it would fail to capture the intricacies of the activity sequence patterns of that demographic group.

4.3.1.2 **Markov Chain Model**

The Markov Chain Model (MCM) or Markov Model is a stochastic model used to model chronologically changing systems usually identified as a Markov process (Bishop, 2006). A Markov process is a special type of discrete-time stochastic process that has the following two assumptions (Mitchell, 1997):

1. The probability distribution of the state at time $t+1$ depends on the state at time t, and does not depend on the previous states leading to the state at time t;
2. A state transition from time t to time $t+1$ is independent of time

In essence, Markov Chain Model is a learning algorithm that learns a Markov Process. A Markov Process is a sequential state transition process where the system changes its state at times t, and the next state of the system depends only on the state right before it and not on the states leading up to it. It is a statistical model and is useful for recognizing temporal patterns. In our study, we view the respondent's activities (reported on the interview) as the *states* of the process and a *transition* as simply doing the next activity (next state). We make an assumption that the respondent's next activity given the activities performed until then depends only on the most recent activity and not on the all the activities leading up to it. While it may seem intuitive to assume that all the previous activities would affect the next activity, due to the complexity involved in truly understanding the full relationship of the activities, we take a simplified approach and make the aforementioned assumption. A simple example of a Markov Chain (left side) and the equivalent *chain* for our model (right side) is illustrated in Figure 40.

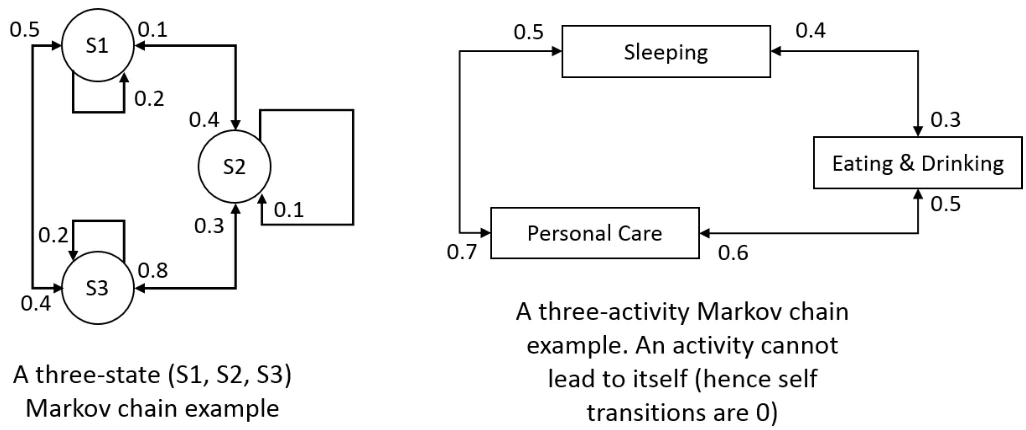

Figure 40 Markov Chains illustrated

In Figure 40, the left side shows a simple three-state Markov chain. The three states S1, S2, and S3 can transition to each other (including themselves) with a probability denoted by the number at the arrow termination. Thus S1 can transition to S1 with a probability of 0.2, to S2 with a probability 0.4 and to S3 with a probability 0.4. On the right side, a simple example using activities are shown. It must be noted that for our activity transitions, we do not consider self-transition (transition to the same state) and hence all self-transitions are assumed to have a probability 0.0; in reality two same activities occurring next to each other would be clumped into one single activity and thus supports our assumption to keep self-transition probability as 0.0.

From the ATUS historical data, each configuration takes the response data from the training year and builds one non-demographic model – a model here thus represents the transition probabilities of each activity to each of the remaining activities. The probability for each transition from the activity to the next activity (read as probability that A_1 occurs before A_2) was calculated as:

$$p(Activity\ A_2|\ Activity\ A_1) = \frac{Number\ of\ times\ A_1\ occured\ immediately\ before\ A_2}{Number\ of\ times\ A_1 occured}$$

It must be noted here that A_1 and A_2 here represent the transformed and coded activities with the temporal ordering A_1 occurred immediately before A_2. For example, if the test configuration number 1 is considered (from **Error! Reference source not found.**), the activities from the ATUS data set are in Tier 3 (T3) code. For each activity (A_1)-next activity (A_2) pair, the first activity (A_1) is transformed to its D-CONCEPT coded activity, while the next activity (A_2) is transformed to its MID coded activity.

Then the response data is divided into the demographic attributes based groups and the demographic models for each group is created. Each of these models (non-demographic and demographic) are then tested against the response data of the testing year. This is performed by taking each activity from each respondent in the testing data set, and using the corresponding trained model to predict the next activity. The actual next activity is then taken and checked to determine if the prediction was correct. This process is repeated for all 60 test configurations for all models.

4.3.1.3 Artificial Neural Networks

Artificial Neural Networks (ANNs) (Yegnanarayana, 2009) are computational methodologies that perform multifactorial analyses. Inspired by networks of biological neurons, artificial neural network models contain layers of simple computing nodes that operate as nonlinear summing devices. These nodes are richly interconnected by weighted connection links, and the weights are adjusted when data are presented to the network during a “training” process. Successful training can result in artificial neural networks that perform tasks such as predicting an output value, classifying an object, approximating a function, recognizing a pattern in multifactorial data, and completing a known pattern. Many applications of artificial neural networks have been reported in the literature, and applications in medicine are growing (Yegnanarayana, 2009). Time series predictions have been conducted with neural networks, including the prediction of irregular and chaotic sequences (Lin et al., 1993, Khashei et al, 2008).

An ANN is usually taken as a black box that accepts a set of inputs and provides one or more outputs depending on the output type. Thus an application of ANN would only see the input and the output nodes, while keeping the functioning hidden. An ANN consists of fundamental ‘processing units’ called neurons connected in a layered arrangement. There are generally three layers in an ANN – an input layer, a hidden layer and an output layer. Each neuron from a layer is connected to every neuron in its next layer and this connection has a weight attached to it. An example of the structure of an ANN is shown in Figure 41.

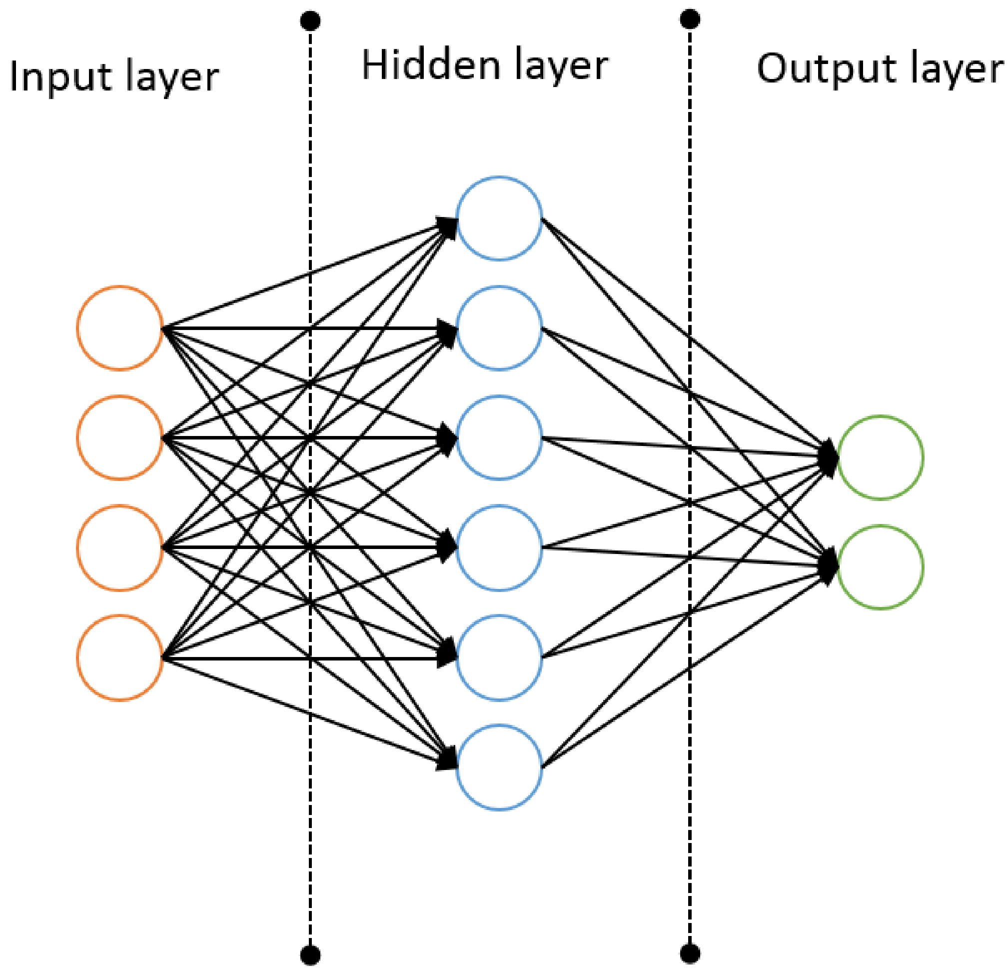

Figure 41 Overview of structure of an Artificial Neural Network

Each neuron thus is connected to all the neurons of the previous layer. The neuron, as the processing unit, is a nonlinear summing node. The input layer neurons are ‘activated’ by the values of the input. They in turn activate the hidden layer neurons which in turn activate the output neurons to provide the output. The structure example of a single perceptron is shown in Figure 42. If S_j denotes the incoming sum for unit j and a_i is the activation value of a unit i, then we have the following:

$$S_j = \sum_{i=0}^{n} w_{ji} a_i$$

$$a_j = \frac{1}{1+e^{-S_j}}$$

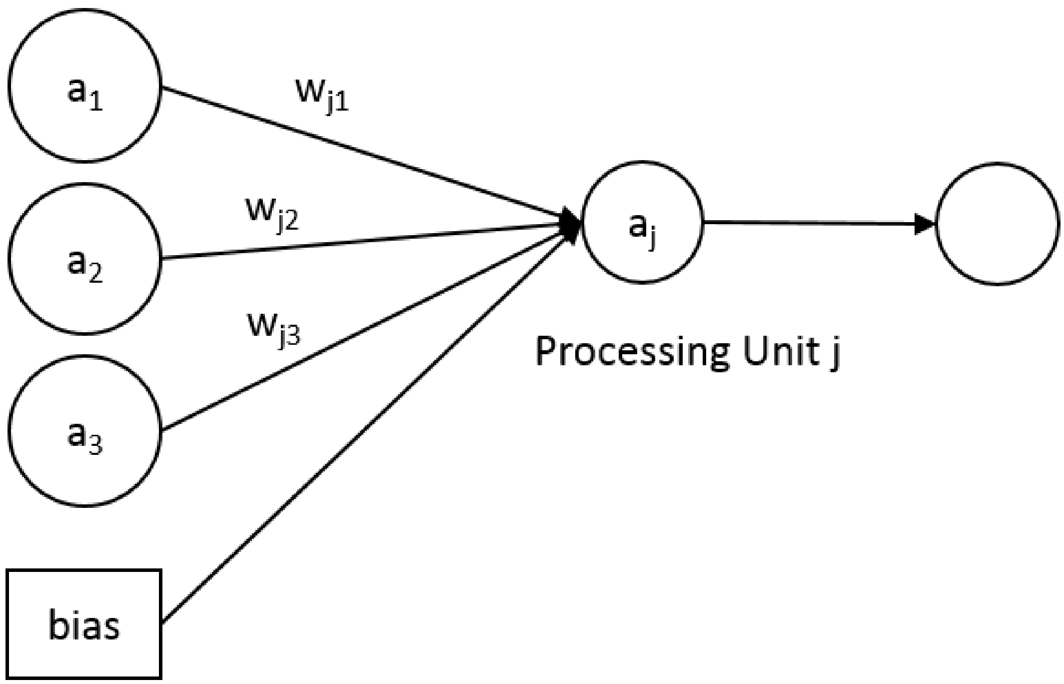

Figure 42 Illustration of an artificial neural network processing unit. Each unit is a nonlinear summing node. The square unit at the bottom is the bias unit, with the activation value set to 1.0. Wji = weight from unit i to unit j

The nodes are generally single class (or binary) with a value of 1.0 for activation state and 0.0 for no activation state. Hence, when multi class values are needed (example, an attribute Gender could have 2 values Male and Female), the classes are split into grouped individual units (in the example, there would be two nodes with Gender=Male and Gender=Female, when the activation node would be the attribute value; if Gender=Male, the Gender=Male would have a value 1.0 while the Gender=Female would have a value 0.0). The output node(s) would also follow a similar pattern based on the type of the output class (single or multi). In our design time Knowledge Engineering, we used the Artificial Neural Network implementation provided with Weka.

4.4 Construction and Deployment

The final stage of the implementation involves putting the system together and deploying it on a publicly accessible server. Our implementation was loaded on to University of Nebraska-Lincoln's Department of Computer Science and Engineering's Intelligent Agents and Multi Agent System (IAMAS) group's lab server. The interviewers from BOSR were trained in using the instrument and on generally conducting time diary surveys using practice interviews. Once the interviewers were ready and the respondents were scheduled, the interviews were conducted by the interviewers from the lab systems at BOSR. For every interview, the audio and screen video was recorded using the recording tool called Camtasia. Post the completion of the phase, the team at BOSR used these recordings to create the transcripts for the interviews.

4.5 Future integration with current setup

As the current implementation is focused on the interviewer-assisted mode (IAM), the system functions with the basic assumptions that the interviewers are motivated users and trained to use the instrument. Our framework is intended to support both IAM and SAM (self-administered mode). The addition of SAM brings about a change in the assumptions about the users: *when the respondents use the instrument directly, they cannot be assumed to be motivated users or to have any significant knowledge about how to the instrument*. One way to handle the lack of knowledge about how to use the instrument is to make the instrument interface intuitive to a new user. But this would only be part of the solution and this section details our ideas for future integration to transition into a true multi-mode setup.

The instructions currently being delivered to the interviewer are static and indicative of the expected action that the interviewers were trained to recognize and use. When the system runs in SAM, the instruction panel can serve as the display outlet for an Interaction Mechanism that could display a personalized message (for example, by leveraging Natural Language Processing) that would better help the respondent in identifying what is expected next similar to the way the interviewer guides the respondent in IAM. This would thus replace the interviewer’s role in assisting the respondent through the survey process.

Another aspect involved in the interface is figuring out what inputs go where. In IAM this would be handled by the trained interviewer and the respondent would only report the information (in many different forms/variations) to the interviewer, who mentally converts the information into the required format. In SAM, this would have to be handled by another Interaction Mechanism backed by a Knowledge Engineering Mechanism to direct the attention of the respondent to the appropriate input field (using highlights or popups). This mechanism would need to have the relevant information as to when to draw the respondent’s attention, what to draw the attention to etc., and this would be provided by the backing Knowledge Engineering Mechanism. This could be accomplished by building models that map accepted sequences of actions and the delays between them.

Furthermore, the current implementation uses a static table lookup for the Prediction KEM and is triggered by the successful submission of an activity. In IAM, this method works fine since the interviewers are motivated users. In SAM however, the users (respondents) cannot be considered to be motivated and thus if they are constantly bombarded with prediction information, unwanted behavior such as satisficing and break-off could happen. Thus the Prediction KEM would have to consider both what is being predicted and when it is being predicted. This can be accomplished by building a set of models of the users and the interviews and using a hybrid recommendation system. In cases where the respondent is highly unmotivated, the system could also attempt to get some data (even if it is satisficed data or bad data) rather than terminating the interview with no data at all. In short, the fidelity of the prediction has to be of a higher level when it comes to SAM.

Thus the process of moving towards SAM from the current implementation of the framework is guided by the observation of the interviewer using the system in IAM and then can be used as the starting point for the respondent using the instrument directly. In Chapter 2, we discussed the systems and methods that are related to our framework by objectives and the processes involved. While Computer Adaptive Testing (CAT) is a very similar knowledge extraction system, it cannot be directly taken into consideration since the users of CAT systems are students and have extremely high motivation and possess the drive to provide as much information as they can provide. Furthermore, it is a system that uses a questionnaire format which is absent in time diary surveys. One of the more significantly related systems is the *Recommendation System* (RS). While they do possess the advantage that the system may (and should) influence the user which is not permitted in time diary surveys, many ideas can be borrowed and adjusted to fit our framework.

For example, according to Pazzani, 2007, Content-Based Recommendation Systems (CBRS) works with associating users with some items according to some lists: an example being web pages in a web search. The problem faced in this is that Natural Language Processing (NLP) is required to handle synonymous and polysemous words. In time diary surveys, this can be redefined as the problem of associating the most appropriate activity sequence prediction knowledge to respondents with acknowledgement of temporal ordering and the interaction history. The respondents may be defined by a set of demographics and the corresponding values, such as the demographic GENDER (coded as PESEX) with values MALE (1) and FEMALE (2). The paper describes how TF-IDF (Term Frequency-Inverse Document Frequency) can be applied to provide a vector space representation of the different words and documents to be recommended. This method however, does not apply a context based weighing since terms such as 'not good' would end up being viewed as two words with separate frequencies rather than negatively identifying it as the opposite of 'good'. The authors provide two existing solutions to bring about this context information and also propose a new method:

1. Using user profiles – where a user profile is a function that predicts the likelihood that the user is interested in an item. This user profile function is based on the history of actions where depending on the domain, certain actions are either avoided or repeated during training; for example, suggesting the same item or

movie is avoided while suggesting a sequel of the movie is encouraged. In our survey system, it would not be possible to create user profiles since the data is de-identified and a repeat respondent cannot be identified. Instead the users can be grouped into population-wide user groups based on their demographics. Preliminary investigations into using the respondent demographics imply that there is variation among the groups, but determining the most suitable combination(s) of demographics was not pursued due to time constraints and lack of sufficient information regarding the users of the system.

2. Manually providing the information usable by the RS – this option simply cannot be used since it places a higher information requirement on the respondent which may lead to break-offs.

3. The author proposed solution is to use a rule-based RS that works on top of the user profiles that provides contextual information regarding the items also. This can be extended as a knowledge-engineered set of rules based on observing the users in action. The knowledge engineering of this can be accomplished by leveraging the data from IAM interviews of similar respondent groups.

Chapter 5: Results

5.1 Introduction

In this chapter we detail the results of our experimental studies and the subsequent data analysis. The experiment was divided into two phases, with each phase having assigned objectives. The objectives of Phase 01 are to determine how well the framework performs as a time diary instrument and the effectiveness of the implemented mechanisms in assisting the interviewer. Section 5.2 describes the experimental setup, process and analysis of Phase 01 in detail. The objective of Phase 02 is to primarily compare and contrast with the data from Phase 01. Phase 02 is also tasked with refining the instrument based on observations from Phase 01. Another objective in Phase 02 is to gather feedback from the interviewers of their evaluation of the framework through the instrument. Section 5.3 deals with the setup for Phase 02 and related analysis of the data.

Once the framework was designed and the implementation completed and tested internally, we, at the Intelligent Agent and Multi Agent Systems (IAMAS) lab collaborated with a team from the Bureau of Sociological Research (BOSR) and Dr. Robert Belli from the Psychology department at the University of Nebraska, Lincoln (UNL) to setup a multi-phase experiment to put the instrument in use to test. We take a phase-wise approach to the experiment so as to enable the knowledge and lessons learned from one phase to be usable in the next. Based on discussions, it was decided to split the experiment into two phases – Phase 01 and Phase 02. Phase 02 is partially complete. A set of interviewers and respondents were recruited for each phase. The interviewers were each assigned a set of respondents who were uniformly sampled from the recruited respondent pool such that every interviewer received the same number of respondents from predefined age groups and genders. The primary limiting factor in the recruitment process was budgetary concerns and not scalability. Thus an interviewer would conduct a specified number of successful interviews for the phase. The interviewers were themselves split into two groups in each of the two phases: one for control and one for treatment. The treatment group is provided with additional instrument feature(s) that would be unavailable to the control group.

5.2 Phase 01

5.2.1 Overview

One of the biggest challenges in evaluating time diary survey data is the absence of the ground truth about the activities reported by the respondent. In time diary surveys, the respondent self-reports the activities that they performed during the diary period. This means that there is no alternate source for verifying if the respondent provided the activities that they actually performed. As a result of the absence of the ground truth, we cannot directly evaluate the quality of the response data collected using our instrument by verifying it against another source. To overcome this problem, we have to create proxy evaluation methods that can provide a means to indirectly verify the quality of the response data. We create proxy evaluation methods by comparing the characteristics of the response data collected against the characteristics of known good quality techniques, and by applying common intuitions and verifying if these intuitions are observable in our response data.

To this end, we collected the data from Phase 01 and divided the analyses of the data into two strategies that employed proxy evaluation methods. The strategies are:

1. First, we establish the quality and a sense of goodness of the response data obtained in Phase 01 by using the American Time Use Survey (ATUS)-a well-known and established as a proxy and comparing our data quality to that reported for ATUS. This also serves to evaluate the effectiveness of the framework in terms of its functioning in an interviewer-assisted mode (IAM). We believe that the response data can be further validated when we proxy intuitions based on how well our predictions are expected to perform at different times of the day. For these, we first report on the quality of the data obtained in Phase 01 in Section 5.2.3 which allows us to determine how well the framework performs as a time diary survey instrument in the IAM mode. Section 5.2.4 then evaluates the predictions made by our instrument by developing co-occurrence matrices to match with the ATUS data. We evaluate the predictions using simple co-occurrence matrices (considers the response data for all interviews), split co-occurrence matrices (considers response data for the interviewer groups separately), equal time co-occurrence matrices (considers the response data divided across equal parts of the day) and primary activity co-occurrence matrices (considers the response data divided by the respondent routine's primary activity). Each of these co-occurrence matrices allows us to understand the prediction matching for each of the identified groups of activities.

2. Once the first strategy is established, our second strategy is to attempt to see how the mechanisms assisted the interviewers during the interview when functioning in IAM. In Section 5.2.5 we first examine the activity creation times at the interviewer level to understand the effect of the predictions and its characteristics on the same. Then, in Section 5.2.6 we discuss the effectiveness of the predictions in assisting the interviewer by studying interview characteristics among the two interviewer groups. Section 5.2.7 then reports on the usage statistics of the different Interaction Mechanisms by the interviewers to understand if they were used effectively or not.

We begin by describing the experimental setup for Phase 01 in Section 5.2.2.

5.2.2 Phase 01 Experimental Setup

The Phase 01 interviews were conducted during June 2015 – July 2015 with four trained interviewers from BOSR. The details of the interviewers are provided in Table 10. Two interviewers were assigned to the control group and two to the treatment group with the treatment group receiving prediction prompts from the instrument which the control group interviewers do not receive. Each interviewer was assigned to conduct 8 successful interviews with an equal distribution of respondents within the three age groups and two genders. For Phase 01, a total of 48 respondents were chosen with each interviewer receiving 12 interviews. Interviewers 24 and 26 received an extra interview each due to having one break-off case each respectively.

Id	Interviewer	Predictions prompted?	Number of interviews (+ breakoffs)
23	Interviewer 23 (I23)	YES	12

24	Interviewer 24 (I24)	NO	12+1
25	Interviewer 25 (I25)	YES	12
26	Interviewer 26 (I26)	NO	12+1

Table 10 Phase 01 interviewer details

The interviewers were trained on how to conduct time diary interviews and in using the instrument. They were provided with multiple practice sessions to get a feel for doing time diary surveys and for using the instrument. The total of the 48 respondents for this phase were equally divided and assigned to the four interviewers. The respondent demographics are detailed in Table 11.

Gender	Age group	Number of respondents
Male	19 – 44	8
Female	19 – 44	8
Male	45 – 64	8
Female	45 – 64	8
Male	65+	8
Female	65+	8
Males: 24; Females: 24	19 – 44: 16;45 – 64: 16;65+: 16	48

Table 11 Phase 01 Respondent demographics details

To reiterate, the purposes of Phase 01 were to:

1. Determine if the framework's instrument implementation performs well as a time diary survey instrument,
2. Study the effects of using the different implemented Interaction and Knowledge Engineering Mechanisms. These include:
    a. The prediction prompts as a separate side panel where it displays the activities that the system predicts would be next. The prediction was done using two methods:
        i. Based on the previous activity (PAB),
        ii. Based on the time of day (TOD).

The prediction prompts were clickable and when a predicted activity is clicked, the corresponding 'activity name' field would be filled with the activity name.

b. The use of the different Interaction Mechanisms to enter activity name and the context fields (who and where). The Interaction Mechanisms available were:

    i. Autocomplete,
    ii. Precode list,
    iii. Manual entry,
    iv. Prediction prompts (for activity name field only).

We hypothesize that:

H1: The interviewer would use the prediction prompt when they feel there is lesser effort involved in clicking the prediction prompt than entering the activity name through other means.

H2: The use of the different data entry methods of the Interaction Mechanisms by an interviewer would increase as they conduct more interviews and become familiar with the instrument.

At the end of Phase 01, the system had collected the response data and the paradata for the 50 interviews conducted. The Camtasia recordings were transcribed by BOSR to produce transcripts for the interviews. Thus after completing Phase 01, we have the response data, paradata, Camtasia recordings, and the interview transcripts. Table 12 lists the different data obtained and their purpose.

Data	Purpose
Response data	Study and analyze the response data to report on the objectives of the framework and their attainment.
Paradata	Study and analyze the paradata recorded during the interviews to report on the interview, interviewer and respondent characteristics.
Camtasia recordings	Used for manual inspection and verification of certain interviewer behaviors.
Interview transcripts	Used for analyzing conversational characteristics of the interviewer and the respondent.

Table 12 Phase 01 data and their purpose overview

5.2.3 Data Quality

As one of the first objective of the framework is to develop an instrument that can be used for conducting time diary surveys, the quality of the data obtained from the use of the instrument needs to be good. While there is no "gold standard" of comparison for time use survey statistics, there are certain data quality measures that have been used in past research (Woods & Wronski, 2013). For ATUS, the metrics used were:

1. Percent of publishable cases: During data editing in ATUS, a small number of cases are removed for one of the two reasons:
    a. If the respondent reports fewer than 5 activities,
    b. If there exists more than 180 minutes of unreported time (refused, gaps).
2. Percent of cases with fewer than 5 activities in the diary,
3. Percent of cases with more than 180 minutes of 'refused/gap' time in the diary,
4. Average number of activities per case.

Older research by Juster (1986) suggests using a very similar set of metrics to measure data quality of time diary surveys consisting of three indicators:

1. The average number of activities per day,
2. The average number of minutes of unspecified time per day
3. The percentage of activities rounded to obvious time slots, e.g., 1 hour or 10 minutes

In our work we take a combined set of the above metrics to understand data quality with respect to our study. They are:

1. (α_1) Average number of activities per interview
2. (α_2) Percent of interviews with fewer than 5 activities and/or with over 180 minutes of unspecified time. Since our framework does not allow time gaps to exist for successful completion of the survey, unspecified time here refers to refusals, don’t know and can’t remember responses.
3. (α_3) Percentage of activities rounded to obvious time slots of 10 and 60 minutes. This rounding is measured based on the way the end time is set. When the stop time is used for denoting the end time of an activity, the minutes of the stop time is checked for rounding while when duration is used, the duration value is used.

The response data from phase 01 was aggregated and the three metrics were calculated. Table 13 details the data quality metrics for our Phase 01 data and also includes the reported values of similar metrics that were available for ATUS, 2013.

Interviewer	Number of interviews	α_1	α_2 (%)	α_3 (%)
I23	12	29.42	0	31.73
I24	13	19.54	7.7	29.13
I25	12	22.92	0	31.64
I26	13	20.00	7.7	35.38
All	50	22.84	4	31.96
ATUS, 2013[1]	38,400	19.6	1.8, 0.5[2]	-

Table 13 Data quality metrics for Phase 01

[1] – As reported by Woods & Wronski, 2013

[2] – This metric for ATUS is reported separately (less than 5 activities and more than 180 minutes of unspecified time)

The average number of activities reported per case in ATUS is 19.6. From Table 13, we see that in Phase 01, the average number of activities recorded per interview (α_1) (equivalent to a case in ATUS) is 22.84. This can be interpreted as the instrument delivering the ability to facilitate the conduction of time diary surveys successfully as the quality of data is similar to the reputable ATUS data quality.

When considering α_2 in Phase 01, though its value of 4% is higher than the reported 1.8% in ATUS for the number of cases that have fewer than 5 activities and/or unspecified time gap more than 180 mins, the number of interviews in Phase 01 is only 50 (including breakoffs). Of these, 2 were breakoff interviews each of which had the remaining of the day that the respondent did not report on marked as a refused activity. Thus of the 50 interviews, 2 interviews failed giving us 4% for α_2. The significance of the 4% however cannot be examined given the small number of interviews.

The ATUS data quality literature did not report on the percent of activities rounded off (α_3), and hence a comparative examination cannot be conducted against Phase 01 results. The reason for selecting α_3 is because rounding off has been associated proportionally with satisficing (Juster, 1986 and Kapelner and Chandler, 2010) and from the point of view of the framework satisficing is an undesirable behavior. From Table 13, α_3 for Phase 01 is about 32%; and since there does not exist any gold standard on the metric, it can be taken as having low satisficing.

One important point to be noted here is the advantages possessed by the interviewers and respondents in ATUS with respect to experience and familiarity when compared to the interviewers and respondents in our phases. The interviewers for ATUS are highly experienced with significant expertise in conducting time diary surveys, while our interviewers were trained for conducting time diary surveys in a brief in-house training session. Furthermore, the respondents in ATUS are pulled from a panel of voluntary participants who would have completed a related survey (ATUS-CPS) a few months beforehand and hence would be well aware of the intents and purposes of ATUS unlike our respondents who were introduced to time diary surveys in a single phone call which also serves to schedule the interview session time. Despite these disadvantages, the response data quality obtained by Phase 01 is comparable to that of ATUS, validating the designs of our instrument to a large extent.

5.2.4 Co-occurrence Matrix Analyses

In this section we present the analysis of the predictions and how they match the data collected with the framework's instrument during Phase 01. We evaluate the matching of the predictions made by our system against the reported activities (response data) by computing the co-occurrence matrices of the predictions and the activities actually entered. The predictions were generated from the ATUS data for the years 2010 through 2013 and thus indicate the occurrences of certain activities in sequence. By relating the predictions on the actual activities recorded, we create an understanding of how our predictions match against the recorded activities and also a map of how our response data relates to the prediction sequences from ATUS.

Here the predictions *p* for an activity *a* are those activities that were recommended as next activities to the interviewer using the prompt panel of the instrument. These predictions include the two types of predictions made using the previous activity based (PAB) and using the time-of-day (TOD). The co-occurrence matrix is calculated by computing the number of times each identified activity was predicted for each actual instance when it was the activity entered by the interviewer. An *identified activity* is a non-verbatim response that the instrument was able to map to an activity in the pre-defined list of activities (auxiliary data). By calculating the co-occurrence matrix, we attempt to understand and interpret the effect of having the prompts (containing the predictions made) on the data collected during the interview, if any.

Each row of the co-occurrence matrix represents an identified activity that was predicted by the instrument, while each column marks an identified activity that was entered by the interviewer. Thus each cell represents the number of times the column activity was actually entered by the interviewer when the row activity was provided as the prediction (either using the previous activity or using the time of day). Understandably this would bring about double counting in the results wherein, an activity predicted by both the methods (using previous activity based and using time of day) would be counted twice for the same actual activity entered. This double counting can serve as a reward/penalty measure since being predicted by both the methods and being the actual activity would get one more count than if it were just predicted by one of the methods. Conversely, if both the methods predicted an activity that was not the actual activity entered, the penalty would be one more count that if were just predicted by one of the methods.

Since the recommendations made by both the methods are visible to the interviewer, the current analysis of the data does not differentiate between which of the methods made the prediction unless otherwise stated. Each method makes 5 predictions primarily, but may only display 4 predictions at times when it removes a prediction that is the activity immediately preceding it. Thus overall, there are between 8 and 10 predictions shown to the interviewer at a time. These predictions are ranked by their decreasing probabilities separately for each method. This ordering is also referred to as the *rank* of the prediction, where the rank is the position it has on the list with rank 1 having the highest probability of occurrence (and hence displayed at the top of the prediction list) and rank 5 having the lowest probability of occurrence (and hence displayed at the bottom of the prediction list).

While generating the co-occurrence matrices the following rules were obeyed:

1. Only primary activities were considered and secondary activities were ignored. Primary activities are those activities that are the ‘main focus’ of the respondent at a point of time while secondary activities are those activities that are done together with a primary activity. For example, if the respondent reports that they were ‘traveling’ while they were ‘talking on the phone’, the ‘traveling’ activity is

recorded as their primary activity while 'talking on the phone' is taken as their secondary activity. ATUS does not record secondary activity information and hence it was taken out of the co-occurrence matrices analysis.

2. The ordering of activities used corresponds to the sequence of activities that were used by the interviewer. Since the interviewer is allowed to switch between activities (and hence bring about previously entered activities to edit later) and predictions were made strictly considering the last activity selected by the interviewer, there are occasions when the predictions made are not meant for the activity currently selected by the interviewer. This occurs when the activity selected by the interviewer is not the activity immediately after it by time. In such situations the predictions and the activity at that point of time are ignored until the interviewer selects an activity in temporal sequence again. Any reference to the ordering of the activities thus refers to the ordering by use of the interviewer rather than the temporal sequence.

   For example, if there are 4 activities in temporal sequence A1, A2, A3, and A4, while the interviewer selects (brings to focus to edit or view) the activities in the sequence A1(a), A2(a), A1(b), A3, A2(b), A4, the predictions made at A1(a) would be considered while the predictions made at A1(b) would be ignored since A3 is not temporally next to A1. Here the bracketed *a* and *b* are *two different instances* of the corresponding activity used to imply that there may have been changes to the activity.

3. If an activity is edited later and newer predictions are made, the predictions made at each instance are considered separately.

For the purpose of understanding if there is a *step offset* between the predictions made and the actual activity, a parameter called *distance* is defined. At *distance d=0*, the predictions made are using the immediately preceding activity. At *distance d=n*, the predictions made are using the activity that was *n+1* positions before it. When referring to the rank used for calculating the co-occurrence matrix, we mean the *cut-off* rank used. Thus the co-occurrence matrix calculated at rank *n* would mean that only the top *n* predictions were considered in computing the matrix.

Since there are 80 identifiable activities in the auxiliary data, the resulting co-occurrence matrix would be of size 80x80. This being a huge matrix makes it hard to both examine and report on and hence, to help better understand co-occurrence matrices, four statistics are computed from each co-occurrence matrix. The statistics computed for the co-occurrence matrix are based off of the work by Soh and Tsatsoulis (1999) and are:

1. Energy, f_1
2. Entropy, f_2
3. Maximum probability, f_3

Along with the above three statistics, a fourth statistic was also included:

4. Sum of diagonal, f_4

The introduction of the co-occurrence matrix is to help understand how the response data collected and the predictions made are related. By computing the four statistics for a co-occurrence matrix, we can develop a simpler way of understanding this relation.

Energy represents how often larger numbers are occurring within the matrix. This translates to how much more often does some predicted activity–actual activity pair occur. If we find high energy in our data, it means that there are instances of predicted activity-actual activity pairs occurring more which means that for some activity the system predicted (rightly) an activity and this occurred frequently, and vice versa. Higher energy thus indicates how often certain activity pairs occurred on both the ATUS data and our response data and indicates higher goodness. While goodness does not have a specific definition in survey literature, here by goodness, we are defining a comparative measure of how well our data matches to the data from ATUS collected in 2013. Thus higher energy indicates that the goodness is similar to that of the ATUS data. Lower energy indicates that the predictions made from the ATUS data do not correlate with the response data obtained and thus indicates that the goodness of the data is not similar to ATUS. High energy is hence a good observation as it indicates that the response data obtained is similar in goodness to that of ATUS, a known good quality time diary survey and having goodness similar to that of ATUS play a part in validating our instrument as a good quality time diary survey too.

Entropy represents how distributed the values are within the matrix. This translates to how much more spread out the predicted activity–actual activity pairs are. The lower the entropy, the more even the spread is. If we find high entropy in our data, it means that there is a higher variance in the activities reported and predicted and these predictions made did not follow the variance correctly, and vice versa. In other words, it means that when the actual activity changes, the predicted activity did not match the change or *follow it*. This indicates that the predictions from the ATUS data and the response data are matching at points of high entropy, but our response data does not vary the same way that the ATUS data. Observing high entropy in the response data is thus not desirable as it is indicating that the response data does not have activity sequences similar to the predictions which are derived from the activity sequences in ATUS.

The *sum of diagonal* represents how high the values along the diagonal of the matrix are. The diagonal of the matrix contains the cells that have the same activity for the row and the column. This translates to how high the occurrences of the same predicted activity and actual activity are. The higher this value, the more the number of times the actual activity is in the predicted activity list. Thus this indicates the *prediction accuracy* of our system, making this measure a particularly important one in our analysis. When higher sum of diagonal is observed, it indicates that our predictions and the activities recorded are matching frequently and is thus highly desirable. Lower sum of diagonal on the other hand, indicates that our predictions did not match the activities recorded, and indicates poor performance by the predictions and is undesirable.

The *maximum probability* represents the maximum probability that was attained for the occurrence of the same actual activity and predicted activity within the normalized co-occurrence matrix. This is a measure of how often was a particular prediction made for an activity regardless of whether it was rightly predicted or not. Lower maximum probability is undesirable, as it indicates that the activities in the response data did not have good predictions made. When the maximum probability is low, it means that the highest probability observed for a predicted activity against an actual activity is low. This means that not enough predictions were made for activities resulting in overall lower probabilities in the matrix and hence a lower maximum probability observed, a behavior that is undesired in our instrument.

The co-occurrence matrices are normalized using the sum of all the values of the matrix. For a normalized matrix P the value at row i and column j is given by $p(i,j)$. The statistics are defined as follows:

$$Energy\ f_1 = \sum_i \sum_j p(i,j)^2$$

$$Entropy\ f_2 = \sum_i \sum_j p(i,j) \log p(i,j)\ where\ \log 0 = 0$$

$$Maximum\ probability\ f_3 = \max_{i,j} p(i,j)$$

$$Sum\ of\ diagonal\ f_4 = \sum_i p(i,i)$$

We then compute four different types of co-occurrence matrices to attempt to examine the effects of different parameters on the predictions. Each type of co-occurrence matrix is differentiated by how the response data is grouped and thus allows us to draw comparisons of the prediction matching across the groups. The first type of co-occurrence matrix, discussed in Section 5.2.4.1, is ***simple co-occurrence matrix*** which is devoid of any grouping and consists of the entire response data from Phase 01; this also helps us determine how our response data is related to the data in ATUS. Then, in Section 5.2.4.2 we discuss the next type of co-occurrence matrix; ***split co-occurrence matrix*** that considers the response data as two separate groups based on if the interviewer conducting the interview were provided the prediction prompts (PROMPT condition) or not (NO-

PROMPT condition). This is then followed by the examination of ***equal time co-occurrence matrix*** in Section 5.2.4.3. In equal time co-occurrence matrix, the response data is divided based on which part of the interview period the activity starts; with the 24-hour interview period divided into six equal parts of four hours starting at 4:00am. Then we examine the ***primary activity co-occurrence matrix*** which divides the response data based on the respondent's reported interview day routine. For each respondent, their reported activities for the interview period are analyzed to determine the 'respondent's primary activity' –the activity that is performed for the longest duration by the respondent. The interview period is then split into three blocks: *the pre-primary block, the primary block* and *the post-primary block*; and the response data is divided based on which block it falls within and analyzed.

5.2.4.1 **Simple Co-occurrence Matrix**

A simple co-occurrence matrix simply considers all of the response data in one go. The ranks used are 1 to 5 and the distances used are 0, 1 and 2. By all the response data, we mean that the data from all the interviews of Phase 01 are considered. Table 14, lists the four statistics computed for the combinations of distances (0 – 2) and ranks (1 – 5) for the simple co-occurrence matrices. Each row in Table 14 gives the row number (shortened to row no.), the distance by which the predictions and the actual activity are offset, the top *n* predictions—i.e., the rank—considered and the four co-occurrence matrix-based statistics: energy, entropy, maximum probability, and sum of diagonal. Again, note that if rank is *n*, then it means that the top *n* predictions were considered for each actual activity in the co-occurrence matrix. This can also be read as at rank *n* for the corresponding distance.

Row no.	Distance	Rank	Energy	Entropy	Maximum probability	Sum of diagonal
1	0	1	0.040	3.952	0.153	0.287
2		2	0.030	4.057	0.101	0.231
3		3	0.022	4.329	0.079	0.188
4		4	0.016	4.570	0.056	0.165
5		5	0.014	4.755	0.043	0.134
6	1	1	0.033	4.013	0.122	0.109
7		2	0.027	4.246	0.120	0.097
8		3	0.024	4.407	0.089	0.090
9		4	0.018	4.608	0.066	0.083
10		5	0.015	4.737	0.053	0.079
11	2	1	0.033	4.051	0.134	0.201
12		2	0.026	4.219	0.089	0.174
13		3	0.021	4.439	0.075	0.151
14		4	0.016	4.671	0.056	0.132
15		5	0.014	4.798	0.049	0.113

Table 14 The four statistics computed for all distance-rank combinations from the simple co-occurrence matrices

From Table 14, we observe at distance 0 where the predictions are for the immediately next activity, the co-occurrence matrix-based statistics perform the best for all ranks i.e. energy, sum of diagonal, maximum probability are maximal while entropy is minimal. So this indicates that our predictions preform the best for the immediately next activity rather than when predicted at greater distances. This implies that activities reported can be approximated to follow a Markov process in that, the next activity is dependent mostly on only the activity that immediately preceded it. This is a desired result, as we assume a Markov process when generating the predictions, and this observation validates the use of predictions based on the previous activity.

To further assist in understanding the data better, the statistics were then grouped by rank and each distance combination was ranked within the group for each statistic. This creates five tables, one for each of the rank. These tables are provided as Table 15 through Table 19. Thus if a distance from the grouped rank combinations (with the same rank) has the best value for a statistic *f*, it gets a score N (or rank 1) for that statistic. This is repeated for each statistic and distance and then a total score is calculated by adding together the score from each statistic. Based on this total score, each row is then ranked further (see the corresponding overall rank column). The grouping by rank enables us to view how the statistics change with increasing number of predictions considered across the distance offsets. Each of the rank group may be referred to as of rank *n*. Thus Table 15 is for rank 1, Table 16 is for rank 2, Table 17 is for rank 3, Table 18 is for rank 4 and Table 19 is for rank 5. Each of the three distances (0, 1 and 2) will have a corresponding row in each of the five tables and thus the corresponding table and row are identified by the rank *n* and the distance *d*. For the tables 15 through 19, each row provides the distance, the rank of the energy, entropy, sum of diagonal and maximum probability for that rank grouped table, the total score based on the ranking of the statistics and the final overall rank. Furthermore, in Table 15 through Table 19, energy, sum of diagonal and max probability are ranked by their decreasing value (higher is desirable) while entropy is ranked by its increasing value (lower is desirable). These values are the raw values as reported in Table 14.

From these tables, we observe that:

1. Distance 0 performs the best compared to all other distances for ranks 1 through 4 based on the overall total rank using the co-occurrence matrix-based statistics. Distance 0 is outperformed by distance 1 in rank 5 (Table 19).
2. The number of statistics that distance 0 is ranked 1decreases as seen in Tables 15 through 19. At rank 1 (Table 15), distance 0 has the best energy rank, entropy rank, sum of diagonal rank and maximum probability rank and hence ranks best in all 4 statistics. At rank 2 (Table 16), distance 0 has the best energy rank, entropy rank and sum of diagonal rank and hence ranks best in 3 of the 4 statistics. At rank 3 (Table 17) and rank 4 (Table 18), distance 0 has the best entropy and sum of diagonal rank and hence ranks best in 2 of the 4 statistics. At rank 5 (Table 19), distance 0 only has the best sum of diagonal, losing out to distance 1 on all other statistics.
3. Another interesting observation is that distance 0 always has the best sum of diagonal—the prediction accuracy of our predictions—value for all the 5 ranks.

Distance	Energy rank	Entropy rank	Sum of diagonal rank	Maximum probability rank	Total score	Overall rank
0	1	1	1	1	12	1
1	3	2	3	3	5	3
2	2	3	2	2	7	2

Table 15 Ranked statistics for the top 1 predictions (simple co-occurrence)

Distance	Energy rank	Entropy rank	Sum of diagonal rank	Maximum probability rank	Total score	Overall rank
0	1	1	1	2	11	1
1	2	3	3	1	7	2
2	3	2	2	3	6	3

Table 16 Ranked statistics for the top 2 predictions (simple co-occurrence)

Distance	Energy rank	Entropy rank	Sum of diagonal rank	Maximum probability rank	Total score	Overall rank
0	2	1	1	2	10	1

1	1	2	3	1	9	2
2	3	3	2	3	5	3

Table 17 Ranked statistics for the top 3 predictions (simple co-occurrence)

Distance	Energy rank	Entropy rank	Sum of diagonal rank	Maximum probability rank	Total score	Overall rank
0	2	1	1	3	9	1
1	1	2	3	1	9	2
2	3	3	2	2	6	3

Table 18 Ranked statistics for the top 4 predictions (simple co-occurrence)

Distance	Energy rank	Entropy rank	Sum of diagonal rank	Maximum probability rank	Total score	Overall rank
0	2	2	1	3	8	2
1	1	1	3	1	10	1
2	3	3	2	2	6	3

Table 19 Ranked statistics for the top 5 predictions (simple co-occurrence)

From observations 1 and 2, we see that as we include a larger set of top *n* predictions, the performance of the predictions for distances greater than 0 begin to match and then surpass the performance of predictions made at distance 0. This could indicate that the top ranked activity is ranked highest because it is the best at predicting the immediately next activity, while the lower ranked predicted activities are ranked lower because they may not be expected as the immediately next activity, but rather they are expected to occur soon, i.e., at distance 1or 2. This could result in the observed trend where distance 0 loses the number of co-occurrence matrix-based statistics in which it performs best as the number of top *n* predictions considered increases, i.e. as the rank increases.

From observation 3, we see that distance 0 predictions have the highest sum of diagonal across all ranks. As the sum of diagonal is a measure of the number of occurrences where the predictions matched the actual recorded activity, it indicates that *the distance 0 predictions are the most accurate when considering the number of* *predictions that match the activity that was actually reported*. This leads to the inference that the predictions made by the system for the immediately next activity are more often right when compared to other pairs of predictions and activities further away in the sequence.

When considering the simple co-occurrence matrix, we can also examine another aspect of the data indirectly. Since the predictions were generated from historical ATUS data, they are a representation of the probability of the top activity sequences that occur within ATUS. The data from ATUS is known to be of good quality and hence if we are able to relate the predictions and the response data, we can attempt to create an indirect measure of the goodness of the data as compared to the ATUS data. This relation will be characterized by high energy (implying the top activities that occur in ATUS also occur as top activities in our data), low entropy (implying that the distribution of the activities in our data is similar to that of ATUS) and high sum of diagonals (implying that the activity sequences from ATUS and the activity sequences from our data are similar). Given that the distance 0 performs well for the top 4 of the 5 ranks, with higher energy, lower entropy, better sum of diagonals and max probability (**Error! Reference source not found.**15 through Table 19), our predictions based on the ATUS data (2010 – 2013) can be said to relate well with the response data obtained from phase 01 at distance 0 and hence the goodness of the data is comparable to that of ATUS.

5.2.4.2 **Split Co-occurrence Matrix**

During the testing in Phase 01, two of the four interviewers were shown the predictions while two interviewers were not shown the predictions. On reviewing the interview videos, we were led to believe that even though the interviewer did not click on the predictions provided (in the prompt panel), they could have still visually used it. To help investigate this, we divide the interviews based on whether the predictions were visible or not into two groups: NO PROMPT (NP) and PROMPT (P). We then compute the co-occurrence matrices separately for the two groups and generate the aggregate and ranked data similar to the process described for the simple co-occurrence matrices. Table 20 lists the co-occurrence matrix-based statistics: energy, entropy, maximum probability and sum of diagonal, for the PROMPT (P) and NO PROMPT (NP) group for each distance and rank.

Distance	Rank	Group	Energy	Entropy	Maximum probability	Sum of diagonal
0	1	P	0.042	3.830	0.148	0.289
		NP	0.042	3.868	0.162	0.283
	2	P	0.029	3.995	0.090	0.218
		NP	0.035	3.923	0.118	0.253
	3	P	0.022	4.254	0.072	0.183
		NP	0.024	4.241	0.092	0.196
	4	P	0.016	4.509	0.050	0.164
		NP	0.018	4.482	0.064	0.165
	5	P	0.013	4.709	0.039	0.133
		NP	0.016	4.671	0.050	0.134
1	1	P	0.038	3.856	0.136	0.096
		NP	0.029	3.950	0.099	0.130
	2	P	0.029	4.145	0.122	0.081
		NP	0.028	4.135	0.117	0.122
	3	P	0.024	4.319	0.085	0.085
		NP	0.025	4.331	0.096	0.099
	4	P	0.018	4.539	0.061	0.080
		NP	0.019	4.528	0.074	0.087
	5	P	0.015	4.684	0.051	0.078
		NP	0.016	4.665	0.057	0.081
2	1	P	0.035	3.926	0.133	0.197
		NP	0.033	3.961	0.135	0.209
	2	P	0.026	4.099	0.084	0.167
		NP	0.027	4.128	0.096	0.186
	3	P	0.022	4.339	0.076	0.153
		NP	0.022	4.367	0.075	0.146

	4	P	0.016	4.604	0.055	0.134
		NP	0.017	4.563	0.060	0.129
	5	P	0.014	4.740	0.047	0.117
		NP	0.016	4.702	0.055	0.115

Table 20 The four statistics computed for all distance-rank combinations from the split co-occurrence matrices

From Table 20, we observe that there appears to be little difference in the co-occurrence matrix-based statistics between the PROMPT and NO PROMPT groups. This indicates that the predictions had *little* effect on the interviewers as they were conducting the interview contrary to our assumption that the interviewer may have visually used it.

Similar to the simple co-occurrence matrix based ranked statistics, the ranking for the split co-occurrence matrices is also computed. This adds another column indicating if the interview data selected is from the PROMPT (P) or the NO PROMPT (NP) group. Table 21 to Table 25 list the co-occurrence matrix-based statistics ranking of each distance, group combination for ranks 1 to 5, respectively. Here, each row in a table for the top *n* predictions (also referred to as at rank n) lists the distance, the group (P or NP), the energy rank, the entropy rank, the sum of diagonal rank, the maximum probability rank, the total score and the overall rank.

Distance	Group	Energy rank	Entropy rank	Sum of diagonal rank	Maximum probability rank	Total score	Overall rank
0	P	2	1	1	2	22	1
0	NP	1	3	2	1	21	2
1	P	3	2	5	3	15	3
1	NP	6	5	5	5	7	6
2	P	4	4	4	5	11	4
2	NP	5	6	3	4	10	5

Table 21 Ranked statistics for the top 1 predictions (split co-occurrence)

Distance	Group	Energy rank	Entropy rank	Sum of diagonal rank	Maximum probability rank	Total score	Overall rank
0	P	3	2	2	5	16	2
0	NP	1	1	1	2	23	1

1	P	2	6	6	1	13	3
1	NP	4	5	5	3	11	5
2	P	6	3	4	6	9	6
2	NP	5	4	3	4	12	4

Table 22 Ranked statistics for the top 2 predictions (split co-occurrence)

Distance	Group	Energy rank	Entropy rank	Sum of diagonal rank	Maximum probability rank	Total score	Overall rank
0	P	4	2	2	6	14	3
0	NP	2	1	1	2	22	1
1	P	3	3	6	3	13	4
1	NP	1	4	5	1	17	2
2	P	6	5	3	4	10	5
2	NP	5	6	4	5	8	6

Table 23 Ranked statistics for the top 3 predictions (split co-occurrence)

Distance	Group	Energy rank	Entropy rank	Sum of diagonal rank	Maximum probability rank	Total score	Overall rank
0	P	5	2	2	6	13	3
0	NP	3	1	1	2	21	1
1	P	2	4	6	3	13	4
1	NP	1	3	5	1	18	2
2	P	6	6	3	5	8	6
2	NP	4	5	4	4	11	5

Table 24 Ranked statistics for the top 4 predictions (split co-occurrence)

Distance	Group	Energy rank	Entropy rank	Sum of diagonal rank	Maximum probability rank	Total score	Overall rank
0	P	6	5	2	6	9	5
0	NP	1	2	1	4	20	1
1	P	3	3	6	3	13	4
1	NP	2	1	5	1	19	2
2	P	5	6	3	5	9	6
2	NP	4	4	4	2	14	3

Table 25 Ranked statistics for the top 5 predictions (split co-occurrence)

1. From Table 21 through Table 25, we observe that distance 0 predictions for both the PROMPT and NO PROMPT group have the highest sum of diagonal statistic rank (1 or 2): which is a measure of the prediction accuracy, across all prediction ranks.
2. We also observe that, when only considering the highest prediction rank (rank 1, Table 21) the PROMPT group performs better than the NO PROMPT group for the

same distance on the overall statistics rank. However, when considering all 5 predictions (rank 5, Table 25), the NO PROMPT group performs better on the overall statistics rank, pushing the PROMPT group to the lower ranks for all distances.

Observation 1 again indicates that the distance 0 predictions were the most accurate as having a higher sum of diagonal rank indicates that the predictions matched the actual reported activities more often, as reported earlier in Section 4.2.1 for simple co-occurrence matrices. This implies that our predictions were most accurate for both the PROMPT and the NO PROMPT interviewers for the immediately next activity.

Observation 2 again indicates that showing the predictions in the prompt panel may have had little effect on the interviewers as the observed changes in the statistics rank between the PROMPT and NO PROMPT groups do not result in an identifiable pattern and could be as a result of random noise in the interviewer behavior. Taking this observation together with the observed lack of difference in the statistics between the groups from Table 20, we can further strengthen the argument that showing the predictions in the prompt panel did not have an effect on the interviewer during the interview. This is not desirable for the instrument and warrants further analysis of how the predictions can best be used to assist the interviewer. This is discussed in Section 5.2.5

5.2.4.3 **Equal Time Co-occurrence Matrix**

From Sections 5.2.4.1 and 5.2.4.2, we can tentatively draw the conclusion that the predictions made by our framework relate well to the response data obtained using our instrument and that it is comparable to the data obtained using ATUS, 2010 - 2013. However, this is only a relative measure and warrants more analysis to understand the quality of data collected and the characteristics of the predictions. We approach this by analyzing the characteristics of the predictions over different times of the day to see if there are any interesting observations across the day. For this, the 24-hour duration from 04:00 am on the day of the interview to 04:00 am the next day is divided into 6 equal time intervals of 4 hours each as follows: 04:00 am – 08:00 am, 08:00 am – 12:00 pm, 12:00 pm– 16:00 pm, 16:00 pm – 20:00 pm, 20:00 pm – 00:00 am and 00:00 am to 04:00 am (next day). The activities in Phase 01 response data are then divided into the corresponding time interval based on the start time of the activity. For example, if an activity starts at 06:45 am, it would be assigned to the 04:00 am – 08:00 am time interval. The co-occurrence matrices are then computed for the activities in each of the time intervals separately and the resulting co-occurrence matrices are called e*qual time co-occurrence matrices*. The four co-occurrence matrix-based statistics are then computed for each equal time co-occurrence matrix. Table 26 lists the rank, distance, time interval group, and the four statistics: energy, entropy, maximum probability and sum of diagonal for each equal time co-occurrence matrix. In the table the group column denotes the corresponding time intervals and it is represented in the format *start time – end time*, where *start time* refers to the starting time of the time interval, and *end time* refers to the ending time of the time interval. For example, the third row lists the statistics for the equal time co-occurrence matrix calculated at rank 1, distance 0 for the activities that fall in the time interval starting at 12:00 pm and ending at 16:00 pm.

Rank	Distance	Group	Energy	Entropy	Max. prob.	Sum of diagonal
1	0	04:00 am - 08:00 am	0.061	3.288	0.153	0.275
		08:00 am - 12:00 pm	0.068	3.303	0.201	0.281
		12:00 pm - 16:00 pm	0.066	3.460	0.203	0.277
		16:00 pm - 20:00 pm	0.054	3.582	0.173	0.267
		20:00 pm - 00:00 am	0.061	3.166	0.117	0.219
		00:00 am - 04:00 am	0.105	2.636	0.232	0.343
2	0	04:00 am - 08:00 am	0.033	3.858	0.082	0.213
		08:00 am - 12:00 pm	0.033	3.960	0.114	0.185
		12:00 pm - 16:00 pm	0.036	4.032	0.120	0.187
		16:00 pm - 20:00 pm	0.035	3.939	0.101	0.205
		20:00 pm - 00:00 am	0.044	3.515	0.117	0.243
		00:00 am - 04:00 am	0.094	2.765	0.205	0.281
3	0	04:00 am - 08:00 am	0.026	4.071	0.069	0.188
		08:00 am - 12:00 pm	0.026	4.167	0.087	0.152
		12:00 pm - 16:00 pm	0.030	4.277	0.097	0.155
		16:00 pm - 20:00 pm	0.026	4.196	0.072	0.178
		20:00 pm - 00:00 am	0.038	3.691	0.082	0.204
		00:00 am - 04:00 am	0.075	2.946	0.156	0.256
4	0	04:00 am - 08:00 am	0.023	4.167	0.060	0.184
		08:00 am - 12:00 pm	0.021	4.341	0.067	0.150
		12:00 pm - 16:00 pm	0.022	4.485	0.077	0.133
		16:00 pm - 20:00 pm	0.020	4.443	0.058	0.144
		20:00 pm - 00:00 am	0.029	3.931	0.069	0.167
		00:00 am - 04:00 am	0.065	3.064	0.121	0.201
5	0	04:00 am - 08:00 am	0.023	4.212	0.055	0.169
		08:00 am - 12:00 pm	0.020	4.410	0.060	0.137
		12:00 pm - 16:00 pm	0.021	4.560	0.068	0.122
		16:00 pm - 20:00 pm	0.020	4.471	0.056	0.137
		20:00 pm - 00:00 am	0.026	4.021	0.063	0.154
		00:00 am - 04:00 am	0.059	3.137	0.111	0.180
1	1	04:00 am - 08:00 am	0.049	3.353	0.103	0.178
		08:00 am - 12:00 pm	0.039	3.700	0.115	0.059
		12:00 pm - 16:00 pm	0.044	3.812	0.166	0.073
		16:00 pm - 20:00 pm	0.036	3.786	0.070	0.106
		20:00 pm - 00:00 am	0.054	3.255	0.132	0.193
		00:00 am - 04:00 am	0.078	2.860	0.177	0.212
2	1	04:00 am - 08:00 am	0.030	3.858	0.080	0.163
		08:00 am - 12:00 pm	0.024	4.169	0.069	0.061

		12:00 pm - 16:00 pm	0.028	4.193	0.085	0.055
		16:00 pm - 20:00 pm	0.026	4.109	0.059	0.106
		20:00 pm - 00:00 am	0.037	3.624	0.093	0.170
		00:00 am - 04:00 am	0.057	3.186	0.133	0.179
3	1	04:00 am - 08:00 am	0.023	4.108	0.061	0.129
		08:00 am - 12:00 pm	0.021	4.349	0.071	0.070
		12:00 pm - 16:00 pm	0.029	4.275	0.111	0.056
		16:00 pm - 20:00 pm	0.024	4.276	0.077	0.106
		20:00 pm - 00:00 am	0.032	3.823	0.069	0.130
		00:00 am - 04:00 am	0.052	3.312	0.116	0.139
4	1	04:00 am - 08:00 am	0.022	4.174	0.063	0.126
		08:00 am - 12:00 pm	0.021	4.413	0.075	0.061
		12:00 pm - 16:00 pm	0.027	4.428	0.086	0.049
		16:00 pm - 20:00 pm	0.021	4.447	0.066	0.092
		20:00 pm - 00:00 am	0.025	4.057	0.056	0.114
		00:00 am - 04:00 am	0.049	3.392	0.092	0.116
5	1	04:00 am - 08:00 am	0.021	4.220	0.059	0.126
		08:00 am - 12:00 pm	0.019	4.480	0.067	0.059
		12:00 pm - 16:00 pm	0.024	4.522	0.077	0.048
		16:00 pm - 20:00 pm	0.020	4.456	0.061	0.091
		20:00 pm - 00:00 am	0.023	4.122	0.054	0.106
		00:00 am - 04:00 am	0.047	3.451	0.100	0.103
1	2	04:00 am - 08:00 am	0.047	3.363	0.115	0.131
		08:00 am - 12:00 pm	0.063	3.492	0.201	0.246
		12:00 pm - 16:00 pm	0.056	3.736	0.188	0.214
		16:00 pm - 20:00 pm	0.045	3.664	0.130	0.167
		20:00 pm - 00:00 am	0.049	3.289	0.089	0.217
		00:00 am - 04:00 am	0.116	2.480	0.193	0.239
2	2	04:00 am - 08:00 am	0.028	3.929	0.073	0.155
		08:00 am - 12:00 pm	0.035	3.975	0.118	0.162
		12:00 pm - 16:00 pm	0.030	4.209	0.106	0.148
		16:00 pm - 20:00 pm	0.029	4.104	0.083	0.154
		20:00 pm - 00:00 am	0.038	3.594	0.071	0.196
		00:00 am - 04:00 am	0.087	2.764	0.177	0.235
3	2	04:00 am - 08:00 am	0.020	4.241	0.059	0.130
		08:00 am - 12:00 pm	0.028	4.199	0.096	0.135
		12:00 pm - 16:00 pm	0.024	4.406	0.081	0.125
		16:00 pm - 20:00 pm	0.025	4.290	0.068	0.146
		20:00 pm - 00:00 am	0.030	3.827	0.065	0.162
		00:00 am - 04:00 am	0.077	2.942	0.138	0.193

4	2	04:00 am - 08:00 am	0.019	4.316	0.050	0.122
		08:00 am - 12:00 pm	0.024	4.294	0.073	0.122
		12:00 pm - 16:00 pm	0.019	4.581	0.064	0.118
		16:00 pm - 20:00 pm	0.020	4.466	0.058	0.123
		20:00 pm - 00:00 am	0.023	4.093	0.051	0.136
		00:00 am - 04:00 am	0.066	3.063	0.128	0.160
5	2	04:00 am - 08:00 am	0.018	4.345	0.052	0.127
		08:00 am - 12:00 pm	0.023	4.370	0.068	0.116
		12:00 pm - 16:00 pm	0.018	4.676	0.060	0.110
		16:00 pm - 20:00 pm	0.019	4.484	0.054	0.120
		20:00 pm - 00:00 am	0.021	4.135	0.046	0.124
		00:00 am - 04:00 am	0.061	3.116	0.121	0.143

Table 26 The four statistics computed for all distance-rank-time-interval combinations from the equal time co-occurrence matrices

One purpose of splitting the day into these time intervals is to understand if certain time intervals have any characteristic predictions. Intuitively thinking, most people would have a much less varying morning schedule from 4 am to 8 am than say later in the evening after 4 pm and this would result in better prediction accuracy during the time interval from 04:00 am to 08:00 am when compared to the prediction accuracy during other time intervals of the day say, 16:00 pm to 20:00 pm. Thus using equal time co-occurrence matrices, we hope to understand if the predictions made during one time interval are better off or worse off than the predictions made for another time interval.

1. From Table 26, we observe that when considering the predictions at distance 0, the time interval from 00:00 am to 04:00 am has the best co-occurrence matrix-based statistics for all ranks 1 through 5, with maximal energy, sum of diagonal and maximum probability and minimal entropy.
2. We also observe in Table 26 that the best sum of diagonal: which is a measure of the prediction accuracy, is at distance 0 for every time interval group across all 5 ranks.

3. Next, at distance 0, when considering the lowest sum of diagonal: which is a measure of the prediction inaccuracy, we observe in Table 26 that except for the top 1 prediction (rank 1), the lowest sum of diagonal is always for the time interval from 12:00 pm to 16:00 pm for the top 2, 3, 4 and 5 predictions (rank 2, rank 3, rank 4 and rank 5). For the top 1 prediction (rank 1), the lowest sum of diagonal is observed to be for the time interval from 20:00 pm to 00:00 am.

Observation 1 indicates that the time interval with the best prediction performance is from 00:00 am to 04:00 am as shown by the observed optimal values for the four co-occurrence matrix-based statistics. This means that we are able to predict the next activity in this time interval more accurately and that the activity sequences during this time interval are better comparable to that in the ATUS data (2010 – 2013) than the other time intervals. It is to be noted here that this time interval is on the day after the respondent's interview day; the 24-hour duration begins at 04:00 am on the respondent's interview day and ends at 04:00 am the day after the respondent's interview day. As this time interval also has the maximal sum of diagonal, we can further infer that it has the best prediction accuracy also since higher sum of diagonal is a measure of the prediction accuracy.

Observation 2 indicates that the best prediction accuracy for every time interval is at distance 0 when considering the top 1 prediction through to the top 5 predictions (rank 1 through 5), as evidenced by the maximal sum of diagonal – which is a measure of the prediction accuracy. This indicates that distance 0 predictions are the most accurate when considering the number of predictions that match the activity that was actually reported throughout the day (every time interval). This ties in with similar observations made in

Section 5.2.4.1 and Section 5.2.4.2 with distance 0 for simple and split co-occurrence matrices respectively.

Next, from observation 3 we see that the lowest sum of diagonal is for the time interval from 12:00 pm to 16:00 pm for the top 2, 3, 4 and 5 predictions. This indicates that during the time interval from 12:00 pm to 16:00 pm, the top 2, top 3, top 4 and top 5 predictions made by the system do not match the actual activity recorded as the sum of diagonal is a measure of the number of occurrences where the predictions matched the actual activity reported. This leads to the inference that the predictions made by the system for the activities that start during the time interval from 12:00 pm to 16:00 pm are not as often right compared to the other time intervals. Additionally, we also see that for the top 1 prediction, the lowest sum of diagonal is for the time interval from 20:00 pm to 00:00 am and can thus infer that top 1 prediction made by the system is more often wrong for activities that start between 20:00 pm and 00:00 am. While it is desirable that our predictions have good prediction accuracy across all the time intervals, there could be time intervals where the respondent's activities are too individualized to be able to be predicted right often. The time interval from 12:00 pm to 16:00 pm is essentially the afternoon hours and the time interval from 20:00 pm to 00:00 am is from night to midnight and both these time intervals may be susceptible to activities that are individualized per respondent. It would thus be of some advantage to personalize the predictions during these time intervals to deal with the lower prediction accuracy in the same. For example, one could resort to case-based predictions instead of statistics-based predictions as an alternative.

Using the equal time co-occurrence matrix, we analyzed the accuracy of our predictions amongst the 6 time intervals of the 24-hour interview period: where the interview period starts at 04:00 am on the respondent's interview day and ends at 04:00 am the day after the respondent's interview day. As the purpose of this analysis was to understand if certain time intervals have better or worse prediction accuracy, based on the observations, we can conclude that our predictions are indeed more accurate at predicting certain time intervals (00:00 am to 04:00 am) and less accurate at predicting certain others (16:00 pm to 20:00 pm and 20:00 pm to 00:00 am). When looking at the time intervals that the predictions are worse off in, we notice that the time intervals 16:00 pm to 20:00 pm and 20:00 pm to 00:00 am can be intuitively thought of as the time intervals where the respondents would have more individualized activity sequences. This leads us to explore an attempt to understand if the respondent's activities during the day and their individuality itself has any effect on the prediction accuracy.

Intuitively thinking, common sense would indicate that most respondents would generally have similar activity sequence routines past midnight when they would be either sleeping or preparing to go to sleep. This intuition, when taken as a proxy, is in line with our observation 1 that, our predictions were most accurate for the time interval from 00:00 am to 04:00 am. Another intuition proxy that is observed is with our lowest prediction accuracy during 12:00 pm to 16:00 pm, when the respondents' routine is likely to be more individualized (observation 3). These proxies provide us with reasoning to strengthen our belief that our predictions are predicting well where they are expected to and performing bad where they are may be expected to.

5.2.4.4 **Primary Activity Co-occurrence Matrix**

From the equal time co-occurrence matrix discussed in Section 5.2.4.3, we were able to look at the performance of the predictions at different time intervals of the day. Our next analysis is then to understand if the respondent's day itself contributed to any characteristics in the prediction performance. For this, we define a respondent's primary activity of the day as the activity done for the longest summed up duration during the day that is not sleeping, eating or personal care activities. We do not consider sleeping, eating and personal care activities as potential primary activities for the respondent as these are general activities that respondents perform on a daily basis and do not necessarily enshrine the respondent individuality that we are concerned with. Once a respondent's primary activity of the day is identified, we break the activities reported in the respondent's day into 3 blocks:

1. Pre-primary: Activities that start between 04:00 am up until the start time of the first occurrence of the respondent's primary activity.
2. Primary: Activities that start in the time interval from the start time of the first occurrence of the primary activity until the stop time of the last occurrence of the respondent's primary activity.
3. Post-primary: Activities that start in the time interval from the stop time of the last occurrence of the respondent's primary activity to 04:00 am the next day.

Following this, we compute the co-occurrence matrices for each block by taking the corresponding activities and predictions from Phase 01 response data that fall within that block for each respondent.

To illustrate this process, consider Table 27; a simplified sample of activities and their start and stop times as reported by a respondent together with the calculated duration:

Activity	Start time	Stop time	Duration (minutes)
Sleeping	04:00 am	06:30 am	150
Personal care	06:30 am	08:30 am	120
Traveling	08:30 am	08:45 am	15
Working	08:45 am	12:45 pm	240
Eating and drinking (not at home)	12:45 pm	13:30 pm	45
Working	13:30 pm	16:30 pm	180
Traveling	16:30 pm	17:15 pm	45
Shopping	17:15 pm	19:00 pm	105
Traveling	19:00 pm	20:00 pm	60
Eating/drinking (home)	20:00 pm	20:30 pm	30
Personal care	20:30 pm	20:45 pm	15
Sleeping	20:45 pm	04:00 am	435

Table 27 Simplified sample of reported activities, start time, stop time and the calculated duration by a respondent

For the sample respondent of Table 27, the respondent's primary activity would be "Working", since the total duration for "Working" is the longest with 240 + 180 = 420 minutes. Note that even though "Sleeping" has a higher total duration, we do *not* consider sleeping, eating or personal care activities for the primary activity as stated earlier, giving us "Working" as the respondent's primary activity. The block assignment for the activities of the sample respondent from Table 27 is listed in Table 28.

Block	Start time	Stop time
Pre-primary	04:00 am	08:45 am
Primary	08:45 am	16:30 pm
Post-primary	16:30 pm	04:00 am

Table 28 The corresponding block assignment for the activities based on the sample respondent's primary activity

Using this approach, we hope to understand if when the reported activities of the respondent's day are divided among these blocks, would there be a block that has its prediction performance better or worse off. Table 29 lists the co-occurrence matrix-based statistics: energy, entropy, maximum probability and sum of diagonal, for each block (under group column) for distances 0 to 2, and rank 1 to 5 (top 1 to top 5 predictions).

Rank	Distance	Group	Energy	Entropy	Maximum probability	Sum of diagonal
1	0	Primary	0.045	3.851	0.166	0.252
		Post-Primary	0.040	3.671	0.101	0.277
		Pre-Primary	0.051	3.661	0.166	0.277
2	0	Primary	0.027	4.236	0.100	0.179
		Post-Primary	0.029	3.998	0.091	0.242
		Pre-Primary	0.030	4.085	0.098	0.204
3	0	Primary	0.022	4.440	0.077	0.153
		Post-Primary	0.024	4.217	0.065	0.203
		Pre-Primary	0.024	4.294	0.080	0.176
4	0	Primary	0.017	4.636	0.059	0.134
		Post-Primary	0.019	4.433	0.051	0.160
		Pre-Primary	0.021	4.399	0.062	0.169
5	0	Primary	0.016	4.675	0.053	0.126
		Post-Primary	0.017	4.483	0.047	0.148
		Pre-Primary	0.020	4.451	0.055	0.153
1	1	Primary	0.030	3.979	0.096	0.069
		Post-Primary	0.032	3.839	0.093	0.172
		Pre-Primary	0.030	3.961	0.087	0.118
2	1	Primary	0.021	4.305	0.062	0.064
		Post-Primary	0.023	4.150	0.064	0.146
		Pre-Primary	0.021	4.284	0.058	0.112
3	1	Primary	0.021	4.435	0.081	0.064
		Post-Primary	0.021	4.313	0.060	0.119
		Pre-Primary	0.020	4.415	0.059	0.107
4	1	Primary	0.019	4.574	0.069	0.056
		Post-Primary	0.018	4.455	0.049	0.099
		Pre-Primary	0.019	4.487	0.067	0.100

5	1	Primary	0.018	4.621	0.063	0.056
		Post-Primary	0.017	4.504	0.044	0.092
		Pre-Primary	0.018	4.553	0.062	0.097
1	2	Primary	0.044	3.863	0.163	0.209
		Post-Primary	0.040	3.694	0.100	0.230
		Pre-Primary	0.038	3.895	0.134	0.168
2	2	Primary	0.026	4.269	0.097	0.154
		Post-Primary	0.027	4.064	0.063	0.200
		Pre-Primary	0.025	4.280	0.092	0.141
3	2	Primary	0.022	4.452	0.077	0.137
		Post-Primary	0.023	4.265	0.054	0.165
		Pre-Primary	0.022	4.458	0.079	0.127
4	2	Primary	0.017	4.607	0.060	0.126
		Post-Primary	0.018	4.462	0.046	0.135
		Pre-Primary	0.019	4.536	0.061	0.115
5	2	Primary	0.017	4.658	0.056	0.120
		Post-Primary	0.017	4.508	0.042	0.124
		Pre-Primary	0.018	4.596	0.060	0.113

Table 29 The four statistics computed for all distance-rank-block combinations from the primary activity co-occurrence matrices

1. From Table 29, we observe that the best sum of diagonal, which is a measure of the prediction accuracy, is for distance 0 across all three blocks and across the top 1 prediction through the top 5 predictions (rank 1 through rank 5).

2. We also observe that the sum of diagonal is the highest for: (a) the pre-primary block and the post-primary block at distance 0 for the top 1 prediction, (b) the post-primary block at distance 0 for the top 2 and top 3 predictions, and (c) the pre-primary block at distance 0 for the top 4 and top 5 predictions. Extending this observation, we also note that at distance 0, for the top 4 and top 5 predictions, the post primary block and pre-primary block have *very similar* sum of diagonal values (i.e., a difference of only 0.009 for the top 4 predictions and 0.005 for the top 5 predictions). Thus the overall observation can be simplified as that, at distance 0, the sum of diagonal is the highest

(or very close to the highest) for the *pre*-primary block and the *post*-primary block for the top 1, top 4 and top 5 predictions, while the sum of diagonal is the highest at distance 0 for the *post*-primary block alone for the top 2 and top 3 predictions.

3. Finally, we also observe that the smallest sum of diagonal, which is a measure of the prediction inaccuracy, is for the primary block for distance 0 and across the top 1 prediction through the top 5 predictions (rank 1 through rank 5).

Observation 1 ties in again with the previously observed best sum of diagonal in simple co-occurrence matrix (Section 5.4.2.1), split co-occurrence matrix (Section 5.4.2.2) and equal time co-occurrence matrix (Section 5.4.2.3) and indicates that for each of the three blocks: pre-primary, primary and post-primary, the predictions are most accurate for the immediately next activity as opposed to the activities further after.

Observation 2 indicates that the top 1, top 4 and top 5 predictions are the most accurate in predicting the immediately next activity that starts in the pre-primary and post-primary block as evidenced by the maximal sum of diagonal at distance 0, since the sum of diagonal is a measure of the number of occurrences where the predictions matched the actual recorded activity. However, for the top 2 and top 3 predictions, the predictions made for the immediately next activity are more accurate for the activities that start in the post-primary block than either of the two other blocks; primary and pre-primary. This allows us to infer that the instrument's predictions were good (where good indicates that the predictions match) for the top 1 through top 5 predictions for the activities that start in the post-primary block, i.e., our predictions for the immediately next activity that starts in the post-primary block are most accurate. Furthermore, for the activities that start in the pre-primary block, the top 1, top 4 and top 5 predictions are accurate. Taken together, observation 2 allows us to infer that making the top 5 predictions during the pre-primary and the post-primary would give us a high prediction accuracy, which is extremely desirable for our instrument.

Finally, observation 3 indicates that the top 1 through top 5 predictions for the immediately next activity that starts in the primary block are more often wrong than right, as evidenced by the minimal sum of diagonal. As a minimal sum of diagonal indicates that the predictions were the least accurate when considering the number of predictions that match the actually reported activity, we can infer that the predictions made by the system for the immediately next activity that starts in the primary block are more often wrong than right when compared to the predictions made for activities starting in the pre-primary or post-primary block. Predicting the wrong activity is not desirable in the instrument, and it can be deemed pertinent that the predictions during the primary block must be more relevant to the respondent based on the respondent's primary activity.

Intuitively, respondents would be preparing to start their day with general routine activity sequences before they begin their primary activity and that, after they are done with their primary activity would return to their residences and then perform their household and personal care activities before sleeping and hence our predictions should be able to predict well before and after the primary activity. These tie in with our observation 2 where our predictions are most accurate for the pre-primary and post-primary blocks. Considering these proxy intuitions, we see further evidence supporting our belief from Section 5.2.4.3, that our predictions are performing well when they can be expected to.

5.2.4.5 **Summary**

Based on the observations from Section 5.2.4.1 through Section 5.2.4.4, we summarize that:

1. The data obtained in Phase 01 through our instrument is comparable to the data obtained through ATUS (2010 – 2013).
2. The predictions made by the instrument are more accurate in predicting the immediately next activity when compared to the prediction accuracy for activities further in the sequence.
3. Showing the predictions through the prompt panel to the interviewer during the interview, however, did *not* fulfil the instrument's purpose of assisting the interviewer as there was no observed difference in the prediction statistics in the data between the PROMPT and NO PROMPT interviews.
4. There are indications that the data collected using our instrument are intuitively correct, based on findings using equal time and primary activity co-occurrence matrices. For example, the predictions made for the immediately next activity, for the activities that start between 00:00 am to 04:00 am are more accurate as compared to the prediction accuracy for activities that start during other time intervals. Also, our predictions made for the immediately next activity, for the activities that start during the pre-primary and post-primary blocks—based on dividing the respondent's day by

their primary activity—are more accurate as compared to the prediction accuracy for activities that start during the primary block.

As discussed earlier, since the concept of the ground truth was unavailable to us to strictly verify and confirm the validity of our predictions, we employed the use of proxies to create a better understanding of our prediction validity and characteristics and were able to observe that our predictions were accurate where expected to, strengthening our support for the validity of the data collected and thus the instrument.

Thus, we complete the first strategy for the analyses of Phase 01 data and have established that the response data obtained in Phase 01 using the IAM implementation of our framework is comparable to that of ATUS (2010 – 2013) and has a sense of goodness. However, our prediction-based analyses did not show that our framework in its IAM mode helped the interviewer noticeably as the PROMPT and NO PROMPT versions did not produce different results. Actually, one could say that our PROMPT version also did not distract the interviewers.

Nevertheless, after analyzing the data, we speculate that, while it was encouraging that the predictions were accurate in many instances, these predictions as prompted could have been rendered unusable due to a particular design issue. More specifically, the design issue of concern is that having the predictions delivered using a prompt panel might not fit within the flow of the interviewers' actions while conducting the interview (the prompt panel was placed to the far right corner of the instrument). Having acknowledged this design issue, an alternate method—thus improving the Interaction

Mechanism—for presenting the predictions was implemented for Phase 02 in an attempt to ensure that the predictions would be of assistance to the interviewer.

5.2.5 Interviewer Characteristics

In this section we study the effect that the predictions had on the interviewer during the interview process. Since the predictions are made when an interviewer submits an activity and creates a new activity (the creation happens automatically and immediately after submitting an activity), one way of identifying if the predictions affected the interviews is to look at the time taken by the interviewer to create and submit an activity. We reason that when predictions are made, it would affect the interviewer when entering the activity information when it's reported by the respondent and thus impact the time taken to create an activity. Thus the time taken by the interviewers to create an activity using our instrument serves as a measure of the data entry time which in turn acts as a proxy for the data collection efficiency of our instrument. The lesser the time taken to create activities, the better the data collection efficiency of our instrument and vice versa.

The time taken by the interviewer to create an activity, also known as the activity creation time, is defined as the time interval from the entry of the first piece of information to the point of time the activity was submitted. This disregards the initial waiting time while communicating with the respondent in certain cases when the interviewer would start the interview before calling up the respondent creating a long waiting time when the first activity is created. This also disregards later edits since it usually involves changes in context information (duration, who, and where) and not the actual activity information and thus the prediction prompts would have no bearing.

Below, in Section 5.2.5.1 we discuss the activity creation times for the two groups of interviewers (PROMPT and NO PROMPT) by comparing them for statistical significance. Then in Section 5.2.5.2, we consider the activity creation times of the PROMPT condition interviewers alone and compare them for statistical significance based on if the predictions that were made matched the actual activity entered by the interviewer. This allows us to examine any effects introduced by having the right predictions which can then be used for improving our instrument in Phase 02.

5.2.5.1 **Interviewers' Activity Creation Times**

In this analysis, we take the activity creation times for all the activities that were predicted for the two interviewer groups. Predictions can be made by one of the two methods: Previous Activity Based (PAB) and Time of Day (TOD). The corresponding data is then split into two sets based on if the interviewers were displayed the prompts (PROMPT) or not (NO PROMPT). An activity is considered to have been predicted when at least one of the prediction methods predicts the activity within its top 5 prediction ranks. Using this we hope to understand if there is a statistically significant difference in the activity creation time between PROMPT and NO PROMPT interviewers when the activity is predicted. When there is a statistical significance in the activity creation times between the PROMPT and NO PROMPT interviewers, the group with the lesser average activity creation time can be considered to have performed better. We consider the predicted activities alone to isolate the effect that making the right predictions would have on the activity creation time.

To determine the statistical significance between the activity creation times of the prompted and not prompted interviewers, we perform the student's t-test. The null hypothesis for the student's t-test here is that there is no difference in the mean of the activity creation times between the prompted and not prompted interviewers for predicted activities. A p-value less than α (=0.05) indicates that the null hypothesis can be rejected and that there is statistical significance in the activity creation times of the predicted activities when they are shown (PROMPT) and not shown (NO PROMPT) to the interviewer.

Table 30 lists the three predicted activity data sources, the two type sets for the student t-test, the number of predicted activities in the data source (Count), the mean activity creation time in seconds and the standard deviation (std. dev) in seconds of the activity creation time. Table 31 then presents the student t-test results between the sets in each data source from Table 30.

Data source	**Type**	**Count**	**Mean (seconds) (± Std. dev)**
Predicted	Prompted interviewers	323	18.07 (± 25.79)
	Not Prompted interviewers	261	17.40 (± 12.16)
Predicted by TOD	Prompted interviewers	288	17.83 (± 26.89)
	Not Prompted interviewers	230	16.99 (± 12.09)
Predicted by PAB	Prompted interviewers	299	18.03 (± 26.00)
	Not Prompted interviewers	241	17.97 (± 12.52)

Table 30 Interviewers' activity creation time count, mean (seconds) and variance for predicted activities when they were prompted and not prompted

Data Source	**df**	**t**	**p**
Predicted	481.29	0.41	0.68
Predicted by TOD	419.11	0.47	0.63
Predicted by PAB	448.07	0.03	0.97

Table 31 Activity creation time for prompted and not prompted interviewers' degree of freedom (df), student's t-test t value and p-value statistics

From the p-values in Table 31, we see that no prediction method has p <= 0.05. Thus we cannot reject the null hypothesis for the t-test and can claim that *there is no statistically significant difference in the predicted activity creation times between prompted and not prompted interviewers*. We also observe from Table 30, that the NO PROMPT interviewers always have a mean activity creation a bit less than that of the PROMPT interviewers and that there are fewer predicted activities for NO PROMPT interviewers than PROMPT interviewers. This difference in average values, however, is not significant as indicated by the standard deviation values. Our current study doesn't provide enough data to understand this difference, though we believe it may not be significant because the PROMPT and NO PROMPT interviewers may be handling certain activities differently though consistently.

Thus we can first conclude that showing the prompts to the interviewers did not create a statistically significant effect on the activity creation times as there is no statistically significant difference between the activity creation times of the PROMPT and NO PROMPT interviewers. Since the PROMPT interviewers did not use the predictions directly (i.e. they did not click the predictions as was intended by design), and given that there is no observable statistical significance in the activity creation times for predicted activity, there could be an issue that the predictions were simply not in an accessible location on the screen for the interviewers; an observation that aligns with that in Section 5.2.4.2. This means that the predictions will have to be delivered through alternate means that would allow them to be used by the interviewer to understand if the predictions have an effect on the interviewer and thus provides us with the opportunity to improve the corresponding Interaction Mechanisms for Phase 02.

5.2.5.2 **Prompted Interviewers' Activity Creation Times**

In this section, we analyze the data from the prompted interviewers alone to understand the effect of the predictions since only the prompted interviewers were shown the predictions. With this analysis we hope to identify if there is any statistically significant difference between the activity creation times for activities that were predicted correctly and those that were not. The prompted interviewers were interviewer 23 (I23) and interviewer 25 (I25). They are called PROMPT interviewers collectively. We discuss the analysis for the PROMPT interviewers by considering whether the activity was predicted correctly or not. When an activity is predicted correctly, it means that at least one of the prediction methods had the actual activity entered by the interviewer within its prediction list that was shown when the activity was created. This considers only the PROMPT interviewers since the NO PROMPT interviewers were not shown the predictions and could not have been affected. This also considers only the predicted activities since we are interested in observing the effect of having the right predictions on the activity creation time.

We consider the activity creation times for only the prompted interviewers and divide the data source based on if the activity that was entered by the interviewer was predicted or not predicted correctly. Table 32 and Table 33 detail the statistics and the student's t-test results for this analysis.

Data source	Type	Count	Mean (s) (± Std. dev.)
Interviewer 23	Predicted	168	19.34 (± 24.26)
	Not Predicted	224	25.99 (± 25.70)
Interviewer 25	Predicted	155	16.69 (± 27.37)
	Not Predicted	141	22.51 (± 17.52)
PROMPT Interviewers (23 and 25)	Predicted	323	18.07 (± 25.79)
	Not Predicted	365	24.65 (± 22.92)

Table 32 Prompted interviewers' activity creation time count, mean (seconds) and variance for activities when they were predicted and not predicted

Data source	df	t	p
Interviewer 23	316.61	-1.10	0.27
Interviewer 25	237.48	-0.47	0.64
Prompt Interviewers	587.28	-1.27	0.20

Table 33 Activity creation time for predicted and not predicted activities for prompted interviewers' degree of freedom (df), student's t-test t value and p-value statistics

From Table 32, we observe that there is no significant patterns in standard deviation between the predicted and not predicted activities for the PROMPT interviewers. We do observe that Interviewer 25 has a relatively smaller standard deviation for activities that were predicted.

From Table 33, we observe that none of the data sources have a p-value less than 0.05. This means that the student t-test's null hypothesis cannot be rejected and that there is no statistically significant difference in the activity creation time for the prompted interviewers when the activity was predicted and not predicted. This observation ties in with the previous observations that the predictions may not be providing the necessary reduction in cognitive load and cements the need to improve the instrument and change the way the predictions are delivered to the interviewers.

5.2.6 Interview Characteristics

In this section we analyze the data at the interview level to understand data collection efficiency and the characteristics of how the predictions affect the data collection efficiency. Through this analysis, we first attempt to measure how efficient our implementation of the framework is in assisting the interviewer collect data during the interview in Section 5.2.6.1, by considering the activities recorded per minute in a session for the PROMPT and NO PROMPT interviewers and using proxies to examine how the predictions affect it. Then in Section 5.2.6.2, we discuss how the session time is affected by the instrument and the characteristic difference between the session time for PROMPT and NO PROMPT interviewers and use a proxy to understand the effect of the predictions in improving the data collection efficiency.

5.2.6.1 **Activities Per Minute Based Analysis**

In this section we examine the average number of activities per minute that was recorded by the interviewers across the two groups: PROMPT and NO PROMPT. The average number of activities per minute recorded by the interviewers serves as a proxy method to understand if the instrument under the prompted and not prompted conditions in IAM affected the interviewers in using the instrument faster thus indicating improved data collection efficiency. The average number of activities per minute is defined as the average of the number of activities recorded per minute in each interview. Thus for PROMPT interviewers, we would calculate the number of activities per minute for each interview that they conducted and then compute the average to obtain the average number of activities per minute. Similarly, we calculate the same for the NO PROMPT interviewers. A higher average number of activities per minute would indicate faster data entry which is desirable for our instrument in IAM as it indicates higher data collection efficiency.

Table 34 lists the average and the standard deviation of the number of activities per minute for the PROMPT and NO PROMPT interviewer groups. Table 35 then lists the student's t-test results for the number of activities per minute of the PROMPT and NO PROMPT interviewer groups.

Interviewer group	Average number of activities per minute (± Std. dev.)
PROMPT	2.001 (± 0.661)
NO PROMPT	1.853 (± 0.498)

Table 34 Average number of activities per minute for each interviewer group

t	1.895
Degree of freedom, df	45.733
p-value	0.064

Table 35 The t value, degree of freedom and p-value for the student's t-test of the average number of activities per minute between the PROMPT and NO PROMPT interviewer groups

From Table 34, again, the standard deviations indicate that there is no significant difference in the spread of the average number of activities per minute between the PROMPT and NO PROMPT group. However, there is an indication that there is an improved usage of the instrument by the PROMPT group over the NO PROMPT group (2.001 vs. 1.853 in terms of average), as the higher average number of activities per minute is indicative of faster data entry which can be considered to be a proxy for the data collection efficiency and hence indicative of instrument usage.

From Table 35, we observe that the p-value for the student's t-test between the number of activities per minute of the PROMPT and NO PROMPT interviewer groups is 0.064 and thus not significant. Thus we cannot state conclusively that PROMPT interviewers were significantly faster than the NO PROMPT interviewers and hence we look at the trends instead for indicative analysis. We thus perform one more analysis with the number of activities per minute against the predictions to examine trends indicative of the performance of the PROMPT interviewers and the NO PROMPT interviewers.

We define the statistic *matched over predicted* as the ratio of the number of predictions that matched the actual activity entered over the number of predictions made in a session. The percent of this can be interpreted as the accuracy of our predictions at the interview (session) level. The value of the matched over predicted percent can range from 0% (where no predictions made matched the actual activity entered) to a maximum value between 20% and 50%. The maximum value varies based on the fact that the most number of predictions that can match the actual activity is at most 2 from the 8 to 10 predictions that the system makes for each activity. As the matched over predicted percent approaches 20% the predictions are more accurate in predicting the actual activity entered. As the matched over predicted percent approaches 20%, the activities entered may be considered to more routine activities since the predictions made consist primarily of routine activities such as eating and drinking, working, traveling etc. We expect that as the respondent reports more routine activities, the interviewers would be able to record them faster. This trend can serve as a proxy to understand if our instrument is able to maintain or increase its effects when making predictions for the PROMPT interviewers. We also generate the regression lines for the two interviewer groups based on a simple linear regression model where the dependent variable is the number of activities per minute in a session and the explanatory variable is the matched over predicted percent. This allows to examine and report on the effect that the matched over predicted percent has on the number of activities per minute in a session in a simple manner.

Figure 43 illustrates the scatter plot for the number of activities per minute for the PROMPT and NO PROMPT interviewers versus the matched over predicted percent together with the corresponding linear regression lines that attempts to fit a simple linear model of the data. The slope, intercept and the standard error for the regression lines of the PROMPT and NO PROMPT interviewer groups in Figure 43 are listed in Table 36.

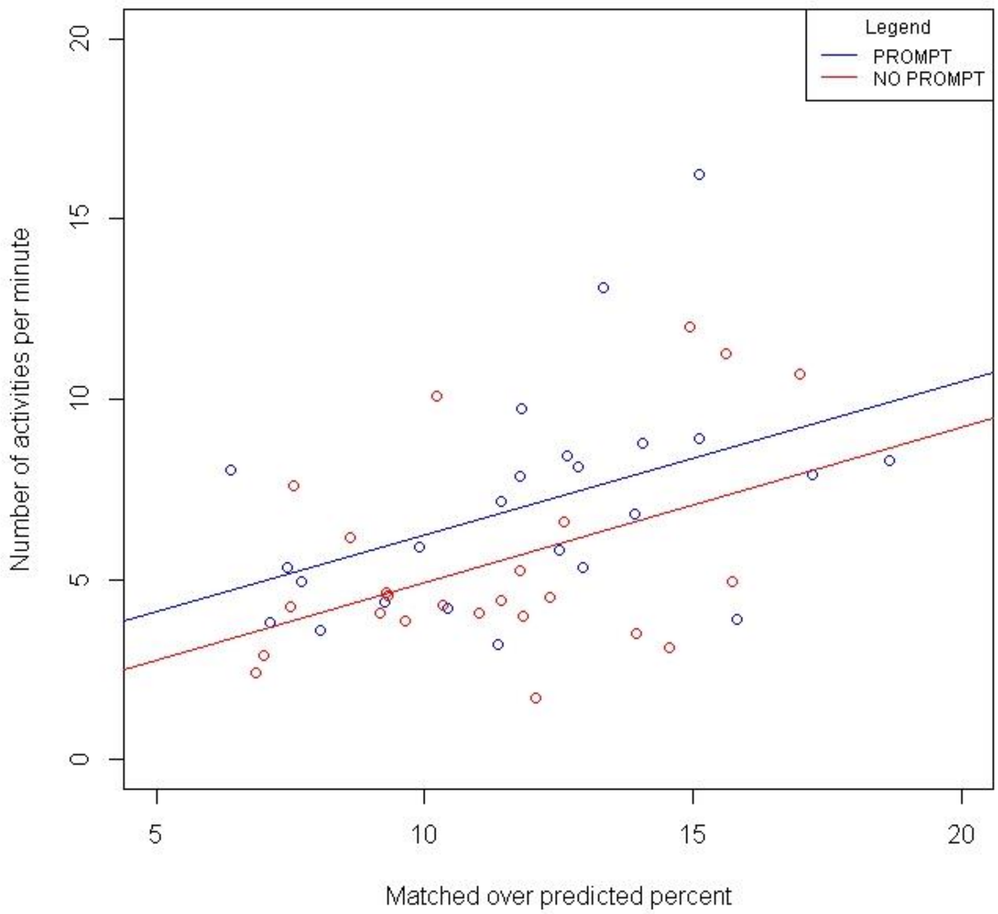

Figure 43 Plots for number of activities per minute versus matched over predicted percent for PROMPT and NO PROMPT interviewers

Interviewer group	Intercept	Slope	Standard error
PROMPT	2.002	0.424	0.180

NO PROMPT	0.582	0.432	0.186

Table 36 Slope and intercepts for the linear regression lines in Figure 43 for the number of activities per minute versus the matched over predicted percent per session for PROMPT and NO PROMPT interviewers

From Table 36, we observe that the standard error for the regression models of the PROMPT and NO PROMPT interviewer groups are close to each other (difference 0.06) which indicates that both the models fit the data in a similar way. This allows us to compare the trends between the two models in an attempt to identify any indicative characteristics.

From Figure 43 and Table 36, we observe that for both the PROMPT and NO PROMPT interviewers, there is a general trend that as the prediction accuracy in the session increases (matched over predicted percent approaches 20%), the number of activities per minute in the session also increases as evidenced by the positive slopes of the linear regression line for both the PROMPT and NO PROMPT interviewer groups. This implies that as the system makes more accurate predictions in a session, the interviewers are able to create activities faster. We also observe that this trend is more pronounced for PROMPT interviewers than NO PROMPT interviewers as evidenced by the larger intercept for the PROMPT interviewer group compared to the NO PROMPT interviewer group and the almost equal slopes of the linear regression lines. Thus we can infer that the instrument exhibits the increased number of activities per minute as the prediction accuracy increases trend as we had expected.

Furthermore, for the PROMPT interviewers, the instrument shows an increased effect for this trend as indicated by the higher intercept for the linear regression line from Table 36. This thus provides with firmer evidence indicating that our PROMPT predictions have an improving effect on the interviewers. This encourages us to believe that with more data from Phase 02, where there is an improved predictions design in the instrument, we would be able to examine the effects of the predictions on the interviewer more closely.

5.2.6.2 Session Time Based Analysis

In this section we examine how the session time varies between the PROMPT and NO PROMPT interviewers. We define the session time as the total time that the interviewer spent in entering data in a session. Table 37 lists the average and standard deviation of the session time in minutes for each of the interviewer groups. Session time cannot be taken directly to imply that one group is better or worse off than the other as the time taken to complete a session depends on the speed with which respondent reports activities together with the interviewer's data entry recording speed. Table 38 then lists the student's t-test results for the session time between the PROMPT and NO PROMPT interviewers.

From Table 37 we observe that the PROMPT interviewers have a higher average—though statistically not significant—session time than the NO PROMPT interviewers. While this observation cannot be directly used to infer a characteristic difference between the two interviewer groups, when this is taken together with the speed of the interviewer group (indicated by the number of activities recorded per minute in a session from Section 5.2.6.1) in recording data we can comment on the characteristics of the interview.

Interviewer group	**Average session time (minutes) (± Std. dev.)**
PROMPT	16.405 (± 10.069)
NO PROMPT	13.200 (± 7.585)

Table 37 Average session time for the interviewer groups

From the observation in Section 5.2.6.1 and from the observation in Table 26, we can state that the PROMPT interviewer group on average records activities faster (higher average number of activities per minute in session) and conducts longer duration sessions (higher average session time). One of the main inferences from these observations is that *the PROMPT group interviewers are recording more activities per session than the NO PROMPT interviewers*. Having more activities recorded per session is a desired outcome of the framework with respect to time diary survey data—implying that such a session is likely to be more precise and thus accurate—and hence this is indicative of an improvement in the performance of the PROMPT interviewers. This is also indicative of better data collection efficiency since having more activities recorded at a faster speed acts as a proxy to improved data collection efficiency—a desired feature for you instrument. However, it is insufficient to make definitive statements or comparisons between the PROMPT and NO PROMPT interviewers and requires more data from Phase 02 to make further conclusions.

t	1.245
Degree of freedom, df	42.746
p-value	0.220

Table 38 The t value, degree of freedom and p-value for the student's t-test of the session time between the PROMPT and NO PROMPT interviewer groups

From Table 38, we observe that the p-value for the student's t-test of the session time between the PROMPT and NO PROMPT interviewers is 0.220. Since the p-value is not less than α (0.05), it implies that the null hypothesis of the student's t-test- that the average session time between the PROMPT and NO PROMPT group is equal- cannot be rejected. This means that there is no statistically significant difference between the session times for the PROMPT and NO PROMPT interviewers. This ties in with the previously examined analysis that we have not been able to observe a statistically significant difference between the PROMPT and NO PROMPT interviewers. Thus we look at indicative trends instead.

With Phase 01 data for the session time, we also attempt to examine how the increasing prediction accuracy—indicated by the matched over predicted percent approaching 20%—affects the session time. We reason that as the prediction accuracy increases, the session time must decrease. This is because, as the prediction accuracy increases, the activities entered as more routine and the interviewers would be able to complete the session faster when there are more routine activities. Unlike the average number of activities per session versus matched over predicted percent, which is taken to indicate how fast the interviewers enter routine activities, the session time versus matched over predicted percent takes on a more interview-wide approach. This trend thus serves as a proxy indicating how much the prediction affects the time taken by the interviewer to complete a session.

Figure 44 illustrates the scatter plot for the session time in minutes versus the matched over predicted percent for the PROMPT and NO PROMPT interviewer per session together with the corresponding linear regression lines. Table 39 then lists the slope and intercept of the linear regression lines from Figure 44 for the PROMPT and NO PROMPT interviewer groups. We also compute and examine the simple linear regression model taking the session time as the dependent variable and the matched over predicted percent as the explanatory variable for simple examination and analysis of the effect the matched over predicted percent has on the session time and for any observable trends.

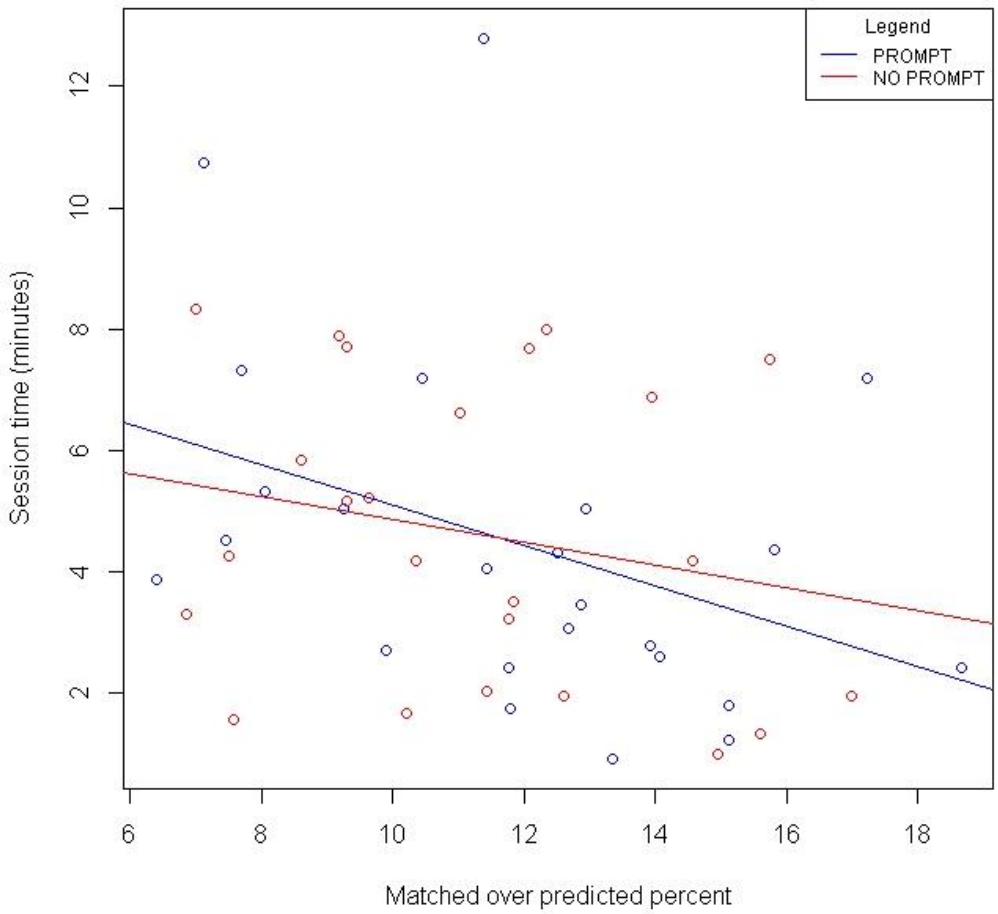

Figure 44 Scatter plots and the corresponding linear regression lines for the session time in minutes versus the matched over predicted percent for PROMPT and NO PROMPT interviewers

Interviewer group	Intercept	Slope	Standard error
PROMPT	8.417	-0.331	0.176
NO PROMPT	6.759	-0.189	0.177

Table 39 The intercept and slope of the linear regression lines of the session time versus the matched over predicted percent for the PROMPT and NO PROMPT interviewer groups

From Figure 44 and Table 39, we observe that as the matched over predicted percent increases, the session time decreases for both the PROMPT and NO PROMPT interviewer groups as shown by the negative slopes for the linear regression lines. From the standard errors reported in Table 39, we observe that the regression models fit the data in the interviewer groups closely as the difference is only 0.001 (0.177 – 0.176 in terms of standard error). We also observe that decreasing session time effect is more pronounced for the PROMPT interviewer group than the NO PROMPT interviewer group as indicated by the steeper slope for the PROMPT interviewer group. It can also be observed that at lower matched over predicted percent (<12%), the NO PROMPT interviewer group has lesser session time than the PROMPT interviewer group and at higher than the 12%, the PROMPT interviewer group has the lesser session time. This means that the PROMPT interviewer group is able to achieve a more pronounced decrease in the session time as the matched over predicted percent increases and becomes better than that of the NO PROMPT interviewers at matched over predicted percent values higher than 12%.

From these observations, we can infer that the both the interviewer groups exhibit the desired and expected trend of decreased session time when there are more routine activities (inferred as the matched over predicted percent approached 20%). The observation of this proxy supports our instrument's objective in enabling faster interviews where expected to. Furthermore, the more accurate predictions allow this trend to be more pronounced and indicates that *having the predictions delivered more suitably can improve our instrument's objective of enabling faster interviews*. This allows us to look forward to the data from Phase 02 where the predictions are delivered through the improved mechanism and thus a stronger improvement should be expected.

5.2.6.3 **Summary**

From the analysis of the data from Phase 01 at an interview level in Section 5.2.6.1 and Section 5.2.6.2, we were able to make the following conclusions:

1. Our instrument is able to indicate that it shows improved data collection efficiency where expected to as evidenced by the increasing number of activities per minute in a session for both the PROMPT and NO PROMPT interviewers. Furthermore, this effect is more pronounced for the PROMPT interviewers where the instrument provides predictions as shown by the higher intercept value and an almost equal slope of the linear regression lines for the number of activities per minute versus the matched over predicted percent when compared to that of the NO PROMPT interviewers. This encourages us to expect an improvement in the data collection efficiency of the instrument when making predictions in IAM with the improved prediction mechanisms implemented in Phase 02.

2. We also observe that the instrument shows results supporting increased data collection efficiency for both the PROMPT and NO PROMPT interviewers as evidenced by the decreasing session time as the matched over predicted percent increases- which again serves as a proxy that the instrument shows improvement in data collection efficiency where expected to. Furthermore, we also observed that this improvement seems to be more pronounced for the PROMPT interviewers than the

NO PROMPT interviewers as shown by the steeper slope for the linear regression line of the session time versus matched over predicted percent which indicates that the system shows improvement in the data collection efficiency where we expect it to. This again encourages us to expect more definitive results from the analysis of Phase 02 where the instrument has been improved and is expected to make the predictions effect the interviewer more.

Thus taken together, we can summarize that our instrument is able to introduce improvement in the data collection efficiency where expected to, which is a highly desirable characteristic and serves as proxies that validate our instrument's objective in improving the data collection efficiency. This improvement can be taken to indicate that the instrument is able to assist the interviewer, however, we are currently unsure on how the assistance is achieved. We also have sufficient information that encourages us to examine the data from Phase 02 where the instrument has improved prediction mechanisms and there would be more data to strengthen the trends that indicate that the instrument effects the interviewer in a positive and desired manner in improving the data collection efficiency.

5.2.7 Entry Method Analysis

In this analysis, we examine the usage of the different data entry methods by the interviewers to understand which method of data entry was preferred by the interviewers. Our instrument implemented two methods for data entry: (1) using *precode* which provides a list of clickable activities above the data entry fields, and (2) using *autocomplete* which provides a list of activities filtered by typing. The third method for data entry is *manual entry*. We are interested in examining how these data entry methods were used for filling the activity name, *who* and *where* data entry fields. For all activities, the activity name is mandatory, while the *who* and *where* fields are never mandatorily required, for some activities (such as sleeping), on the other hand, they are mandatorily not required/allowed. Using this analysis, we hope to understand if the interviewers were able to use the different data entry methods and if they preferred one method over another. This allows us to investigate the usefulness of our Interaction Mechanisms as part of the integrated framework.

5.2.7.1 **Entry Method Percent Analysis**

Table 40 lists the percent of the number of times the precode, autocomplete and manual data entry methods were used to enter the data in the three data entry fields (Activity name, who and where) for the PROMPT, NO PROMPT and ALL interviewers.

Data entry field	**Interviewer group**	**Precode**	**Autocomplete**	**Manual**
Activity name	PROMPT	80.26	3.48	16.26
	NO PROMPT	52.21	31.46	16.33
	ALL	67.35	16.37	16.29
Who	PROMPT	86.03	0.93	13.04
	NO PROMPT	81.18	3.63	15.19
	ALL	83.84	2.15	14.01
Where	PROMPT	97.86	0	2.14
	NO PROMPT	89.80	3.33	6.87
	ALL	94.26	1.48	4.25

Table 40 Percent of number of times each data entry method was used

From Table 40, we observe that the data entry method using precode has the highest percent of usage for all three data entry fields, with 67.35% for activity name, 83.84% for who, and 94.26% for where, taking all the interviewers. We also observe that each interviewer group individually also has the precode as their data entry method with the highest percent of usage. This means that, the interviewers would more often prefer to use the precode to fill the data entry fields, rather than manually entering the data or using the autocomplete. The precode was designed to allow the interviewer to easily click on it to fill in the corresponding entry field, thus by nature, making it the fastest way to enter the data when compared to manually entering it or using autocomplete. Thus, we can infer that the *precode data entry method was the most favored method to enter data* by the interviewers. Since the precode and the autocomplete combined has a higher percent of usage than manual entry, we can further state that the interviewers mostly favored the option to not have to manually type the data in. This inference allows us to validate the logic of having Interaction Mechanisms such as the precode to provide the interviewer with an alternate option to manually typing the data. This provides us with the evidence to further validate the usefulness of the instrument in assisting the interviewers conduct time diary surveys.

Another interesting observation from Table 40 is the relatively large difference in the precode usage percent between the PROMPT and NO PROMPT interviewers for the activity name entry field which is 28.05% (80.26% for PROMPT, 52.21% for NO PROMPT). This large difference is not observed between the two groups for the who field (difference is 4.85%) or the where field (difference is 8.06). However, the interpretation of this observation is not significant as the observation could have been as a result of an individual interviewer's characteristics. Table 41 displays the precode usage percent for each interviewer separately for the activity name entry field.

Interviewer group	Interviewer	Precode percent for activity name
PROMPT	I23	77.10

	I25	84.46
NO PROMPT	I24	22.74
	I26	82.70

Table 41 Individual interviewers' precode usage percent for activity name entry field

From Table 41, we observe that in the NO PROMPT interviewer group, interviewer 24 (I24) has a drastically lower precode usage percent (22.74%) as opposed to the other three interviewers (all greater than 75%). Since interviewer 24 is a NO PROMPT interviewer and this low usage was not shown by the other NO PROMPT interviewer, and given the limited data we have, the difference between the PROMPT and NO PROMPT groups' difference in the precode usage percent for the activity name entry field could be possibly explained as due to interviewer 24's characteristic behavior of low precode usage.

Thus, through this analysis, we can state that the interviewers were able to use the implemented Interaction Mechanisms for data entry well and in particular favored the Precode Interaction Mechanism to enter data faster. This could also potentially explain how our instrument was able to achieve data quality comparable to that of ATUS even though the ATUS interviewers would have had far more experience and training in conducting interviewers, as the interviewers were able to leverage our instrument's Interaction Mechanisms to compensate.

5.2.7.2 **Activity Creation Time with Prediction and Precode Analysis**

In this section, we probe the data from the PROMPT interviewers to understand if using the precodes when the predictions were made correctly influenced the interviewers. We are interested in this analysis since the precodes consists of the same set of activities that are used in the predictions. Furthermore, from observation 3 for Table 40, we identified a more noticeable difference in the precode usage for activity name by the PROMPT interviewers as opposed to the NO PROMPT interviewers for the other data entry fields. Since the PROMPT interviewers who were shown the predictions did not click on it at all, we would like to know if they clicked on the precode by visually processing the predictions. Though we were not able to identify any significant differences previously in any overall analysis; in this analysis, we look at only the PROMPT interviewers and examine the difference in the activity creation time of activities where the predictions matched the actual activity and the interviewer used the precode. Table 42 displays the student's t-test results between the activity creation times of those activities that were predicted and whose activity name were filled using precode and those activities that were either not predicted or whose activity name was not filled using precode for PROMPT interviewers.

t	-6.164
df	655.150
p-value	1.237e-09
Mean activity creation time of activities that were predicted and filled using precode	15.773 seconds
Mean activity creation time of activities that were either not predicted or not filled using precode	25.338 seconds

Table 42 Student's t-test result between activity creation times of activities that were predicted and filled using precode and activities that were not predicted or filled using precode for PROMPT interviewers

From Table 42, we observe that the p-value of the student's t-test is 1.237e-09, which is less than 0.05. Thus the null hypothesis—that the means of the two tested sets are equal—can be rejected and we can state that *there is a statistically significant difference* in the activity creation time of activities that were predicted and filled using precodes and those that were not. This means that *the PROMPT interviewers were able to create an activity that was predicted by using the precode faster than when not using the* *precodes when predicted or when not predicted*. This allows us to infer that the predictions may have actually been visually processed by the interviewer and when the interviewers used the precode they were able to improve (decrease) their activity creation times. This improvement is established as observed from the lower mean activity creation time (15.773 seconds) for the activities that were predicted and filled using precode as compared to the alternate set (25.338 seconds) (note: a lower activity creation time means the activity was created faster, which is a positive observation).

Thus, on observing and understanding that the precodes were used more frequently than the other data entry mechanisms, we were able to strengthen our evaluation that the Interaction Mechanisms (and thus the instrument and the framework) were useful in assisting the interviewer and was able to help the framework attain its objective. We also observed that the activities for which the predictions were right and that were filled using the precode for the activity name were created faster by the interviewers; this provides us with sufficient information to update the instrument for Phase 02 wherein we attempt to leverage the use of precode to deliver the predictions to the interviewer. We do this by modifying the predictions delivery in Phase 02 to be done through the precode mechanism and analyze the effectiveness of this change in Section 5.3.5.

5.2.8 Summary

From the analysis of the data collected from Phase 01 of our experiment, we were able to demonstrate the positive qualities of our instrument and the overlaying framework in attaining the phase objectives. In Section 5.2.3, we were able to show the data quality and the goodness of the instrument when compared to ATUS, 2010 – a known good quality time diary survey instrument. We were then able to use proxies – due to lack of ground truths in time diary surveys – to understand how the Prediction Knowledge Engineering mechanism makes accurate and timely predictions that could be of use to an interviewer in IAM in Section 5.2.4. We were also able to comprehend that the predictions were less useful than expected due to the design flaw that the predictions were not easily accessible to the interviewers; an attempt to fix this design flaw was implemented for Phase 02. With Section 5.2.5, we observed that the predictions did not affect the interviewer's activity creation times – which further adds to the observation that the predictions did not affect the interviewers where expected. Section 5.2.6 then allowed us to understand that the PROMPT interviewers were generally better at using the instrument than the NO PROMPT interviewers as was shown by the activities recorded per minute. This observation was further supported by the difference in the trends observed in the performance between the PROMPT and NO PROMPT interviewers.

5.3 Phase 02

5.3.1 Overview

Based on a preliminary analysis of the data obtained from Phase 01, Phase 02 was planned and began in November 2015. Four new interviewers were selected (due to unavailability of the Phase 01 interviewers) for Phase 02 and the instrument's design of the Prediction Mechanism was modified to allow the predictions to be delivered through the Precode Mechanism. Furthermore, to better understand the effectiveness of the instrument with respect to the interviewers themselves, a questionnaire survey was presented to them on each interview completion. This survey, termed Post Interview Survey, was filled out by the interviewer and thus is representative of the interviewers' feedback regarding different aspects of the interview.

The objectives of the instrument for Phase 02 were:

1. To continue performing as a time diary survey instrument in interviewer-assisted mode (IAM). We analyze the data quality of Phase 02 data in Section 5.3.3 to understand this.

2. Understand the usage of the improved Prediction Mechanisms in Phase 02 wherein, the predictions are delivered through the Precode Mechanism. For this, we investigate for differences in performance between the PROMPT and NO PROMPT interviewers in Phase 02 relating to the predictions in Section 5.3.4. This also demonstrates that mechanisms in the framework can be modified and parts of their working switched to fulfil change in circumstances/requirements.

3. Obtain direct feedback from the interviewers in IAM to gather information regarding their opinion on the usefulness and the impact of the instrument. This feedback allows us to examine the instrument's working as a time diary survey instrument in IAM from the viewpoint of the interviewers itself to provide support for objectives 1 and 2. This analysis is presented in Section 5.3.5.

5.3.2 Phase 02 Experimental Setup

The Phase 02 interviews began in November 2015 and continues through February 2016 and is pending completion. The data used in these analyses are thus limited to those interviews completed before March 1, 2016. Similar to Phase 01, four interviewers were divided into two groups – control and treatment, where the treatment group received predictions during the interview. While Phase 02 is intended to have 48 completed interviews, only 31 interviews had reached completion at the point of writing of this analysis. The interview distribution for each interviewer in Phase 02 (both intended and current) is provided in Table 43. Following this, Table 44 lists the interview distribution for the respondent groups across the interviewers.

Id	Interviewer	Predictions prompted	Number of interviews (intended)	Number of interviews (completed)
28	Interviewer 28 (I28)	YES	12	10
31	Interviewer 31 (I31)	YES	12	6
29	Interviewer 29 (I29)	NO	12	10
30	Interviewer 30 (I30)	NO	12	5

Table 43 Phase 02 interviewer details (16 total interviews total for the PROMPT condition, and 15 for the NO PROMPT condition)

Gender	Age group	Total (current)	I28	I29	I30	I31
Male	19 - 44	8 (3)	2 (1)	2 (2)	2 (0)	2 (0)
Male	45 - 64	8 (3)	2 (2)	2 (1)	2 (0)	2 (0)
Male	65+	8 (5)	2 (1)	2 (2)	2 (0)	2 (2)
Female	19 - 44	8 (6)	2 (2)	2 (2)	2 (1)	2 (1)
Female	45 - 64	8 (7)	2 (2)	2 (1)	2 (2)	2 (2)
Female	65+	8 (7)	2 (2)	2 (2)	2 (2)	2 (1)
Male: 24; Female: 24	19 - 44: 16; 45 - 64: 16; 65+: 16	48 (31)	12 (10)	12 (10)	12 (5)	12 (6)

Table 44 Phase 02 interviews distribution for respondent groups across the interviewers. Note: Read as intended count (current count) for columns 3 through 7

Some of the most important changes in the instrument in Phase 02 from that of Phase 01 are:

1. Predictions are made only using Previous Activity Based (PAB).

2. Predictions are delivered through the Precodes.

3. Administration of the Post Interview Survey for interviewers.

Through the analysis of the available data from Phase 02, we hope to understand whether the changes made to the system were effective in keeping the instrument's purpose of delivering time diary surveys to interviewers (IAM). For this, we study the data quality of the response data collected and investigate for performance difference between the PROMPT and NO PROMPT interviewers with respect to interview characteristics and the predictions. Finally, we also discuss the post-interview survey response submitted by the interviewers to support the observations made previously based on the data.

5.3.3 Data Quality

Similar to the data quality analysis we discussed earlier for Phase 01 data in Section 5.2.3, in this section, we present and analyze the data quality of the Phase 02 data. To reiterate, we consider the following metrics for the data quality of the time diary survey:

4. (α_1) Average number of activities per interview

5. (α_2) Percent of interviews with fewer than 5 activities and/or with over 180 minutes of unspecified time. Since our framework does not allow time gaps to exist for successful completion of the survey, unspecified time here refers to refusals: don't know and can't remember responses.

6. (α_3) Percentage of activities rounded to obvious time slots of 10 and 60 minutes. This rounding is measured based on the way the end time is set. When the stop time is used for denoting the end time of an activity, the minutes of the stop time is checked for rounding while when duration is used, the duration value is used.

We then compare these data quality metrics for the current Phase 02 data with those of ATUS, 2013 (a known good quality time diary survey) metrics and with Phase 01 data quality metrics to understand how the instrument performed as a time diary survey instrument following the modifications that the instrument underwent for Phase 02. Table 45 lists the three data quality metrics for Phase 02 data, Phase 01 data and the reported values from ATUS, 2013.

Interviewer (Type)	**Number of interviews**	**α_1**	**α_2 (%)**	**α_3 (%)**
I28 (PROMPT)	10	23.60	0	33.62
I31 (PROMPT)	6	23.83	0	36.17
I29 (NO PROMPT)	10	24.70	0	36.32
I30 (NO PROMPT)	5	23.40	0	38.60
Phase 02 All	31	23.96	0	35.80
Phase 01 All	50	22.84	4	31.96
ATUS, 2013[1]	38,400	19.6	1.8, 0.5[2]	-

Table 45 Data quality metrics for the interviewers of Phase 02 and the data quality metrics for Phase 01 data and those reported for ATUS, 2013

[1] – As reported by Woods & Wronski, 2013
[2] – This metric for ATUS is reported separately (less than 5 activities and more than 180 minutes of unspecified time)

From Table 45, we observe that the α_1 metric for Phase 02 data is 23.96. This means that the Phase 02 interviews had, on average, 23.96 activities recorded per interview. This is slightly higher (1.12 activities per interview) than the value that was observed for Phase 01 data; which in turn was higher than that reported for ATUS, 2013. Thus, this means that the Phase 02 instrument was able to provide a small improvement in the quality of the data collected using it in IAM and hence the data quality indicators were better. Furthermore, this also means that the instrument did not lose out on its effectiveness as a time diary survey instrument in Phase 02. This serves to provide support that the modifications that were performed on the instrument based on the analysis of the Phase 01 data did not affect the working of the instrument as a time diary survey instrument in a negative manner – wherein, the instrument was able to perform just as well as it did in Phase 01.

There were no break-off interviews in Phase 02 and thus α_2 metric for Phase 02 data is 0. Given, that not all of Phase 02 interviews have been conducted, no significant information can be drawn for this observation. The α_3 metric for Phase 02 is 35.80% - this means that the activity durations/end times in the response data in Phase 02 was rounded-off around 36% of the time. From Table 45, we recall that α_3 for Phase 01 was close to 32%, which is approximately 4% less than that was observed for Phase 02. α_3 serves as an indicator for satisficing in the data –a high value is undesirable, as it indicates lower data quality. Since there exists no gold standard for the value of α_3, it can again be taken as an indication for low satisficing (similar to Phase 01)—which is a desirable trait in our instrument.

5.3.4 Interviewer Characteristics

In this section, we examine the data from Phase 02 at the interviewer level to understand if there is a difference in the time taken for the PROMPT and NO PROMPT interviewers to create activities. Remember that, from Section 5.2, the time taken to create activities is indicative of how fast the interviewers are able to extract the required information from the respondents and record the information using our instrument. To understand the effect of the modification to the Prediction Mechanisms in Phase 02, we narrow down the activities we examine to those that were correctly predicted by our instrument. While our instrument predicts the next activity for both the PROMPT and NO PROMPT interviewers, the predictions are only delivered to the PROMPT interviewers (through the Precode Interaction Mechanism). With this, we hope to understand if the PROMPT and NO PROMPT interviewers had any difference in the way they extracted and recorded those activities that were predicted correctly by the system. Table 46 presents the mean activity creation time (in seconds) and the standard deviation (Std. dev) for those activities that were correctly predicted by the instrument for the PROMPT and NO PROMPT interviewers of Phase 01 and Phase 02. Table 49 then provides the results of the student's t-test for examining the statistical significance between the PROMPT and NO PROMPT interviewers of Phase 01 and Phase 02.

Data Source	Interviewer group	Number of predicted activities per interview	Mean (seconds) (± Std. dev)
Phase 01	PROMPT	13.46	18.07 (± 25.79)
	NO PROMPT	10.87	17.40 (± 12.16)
Phase 02	PROMPT	11.94	16.26 (± 15.08)
	NO PROMPT	12.00	26.19 (± 29.94)

Table 46 Activity creation time statistics for predicted activities between PROMPT and NO PROMPT interviewers in Phase 01 and Phase 02

Data Source (Phase)	**df**	**t**	**p**
Predicted activities (01)	481.29	0.41	0.68
Predicted activities (02)	260.76	-3.99	8.39e-05

Table 47 Student's t-test results for the activity creation time of predicted activities between the PROMPT and NO PROMPT interviewers in Phase 01 and Phase 02 respectively. (df = degree of freedom)

From Table 46, we observe that the mean activity creation time for PROMPT interviewers in Phase 02 is 9.93 seconds less than that of the NO PROMPT interviewers in Phase 02. This means that the PROMPT interviewers in Phase 02 are able to create activities that the instrument predicts correctly faster than the NO PROMPT interviewers. This can be explained by the fact that the PROMPT interviewers in Phase 02 are shown the predictions for the activity (by highlighting the precodes) while NO PROMPT interviewers are not, and thus the PROMPT interviewers are able to easily identify and enter the activity in the instrument. This is an observation that is very encouraging since it provides us with evidence that: the predictions being delivered to the interviewers through the precodes are able to reduce the time taken by the PROMPT interviewers to create those activities. This hints at a result of the reduced cognitive load on the interviewers as they do not have to visually process and search the precodes for the activity. Remember that the predictions are highlighted distinctly in yellow color in the precodes and the interviewers are able to employ this distinction to quickly select the corresponding activity precode. This observation is further strengthened by the student's t-test result in Table 47 which shows a *p* value very close to 0 (8.39e-05) which is less than α (=0.05). This means that there is a statistically significant difference in the activity creation times of those activities that were correctly predicted by our instrument between the PROMPT and NO PROMPT interviewers. Furthermore, from Table 47, we also recall that this difference was not observed between the PROMPT and NO PROMPT interviewers in Phase 01 (*p* value 0.68 > 0.05); we had attributed this to the design issue where the predictions were delivered through a separate panel that the interviewers choose not to use (possibly due to being placed to the far right of the instrument). This statistically significant difference in the activity creation times of predicted activities between the PROMPT and NO PROMPT interviewers can be due to the improved prediction delivery mechanism that was implemented in Phase 02.

However, one could argue that this difference in the activity creation times for the predicted activities could be due to the types of respondents in Phase 02 reporting activities that were easier to record as compared to those from Phase 01. Since the activities that the instrument predicts are generally common activities (such as sleeping, eating, etc.), the argument can be stated that the respondents in Phase 02 would have been easier to interview than those from Phase 01 if the Phase 02 respondents reported more general activities that the instrument predicts. To test this argument, we examine the *average prediction accuracy per interview* for the interviewer groups between the two phases. The average prediction accuracy per interview serves to indicate how much of the activities that respondents reported in an interview were general or common activities, as the prediction accuracy would increase if the respondents report more general activities. Table 48 lists the average prediction accuracy per interview for the two interviewer groups for Phase 01 and Phase 02.

Phase	**Interviewer group**	**Average prediction accuracy per interview (%)**
01	PROMPT	51.78
	NO PROMPT	47.86
	All	49.74
02	PROMPT	50.74

	NO PROMPT	51.76
	All	51.13

Table 48 Average prediction accuracy per interview for PROMPT and NO PROMPT interviewers in Phase 01 and Phase 02

From Table 48, we observe that the value difference in the average prediction accuracy per interview between the PROMPT interviewers for Phase 01 and Phase 02 is 1.04% (51.78 – 50.74), between NO PROMPT interviewers is 3.90% (47.86 – 51.76) and 1.39% (49.74 – 51.13) when taking all the four interviewers of each phase. Since these differences are not significantly large, it can be said that the respondents of Phase 01 and Phase 02 were not different in how easy or hard they were to interview based on their recorded activities.

Thus, we find supporting evidence to strengthen our observation that the PROMPT interviewers were able to record the activities that the instrument predicts correctly faster (9.93s average) due to the predictions being delivered through the precodes. This, thus justifies our reasoning to modify and improve the Prediction Mechanisms to deliver the predictions through the Precode Interaction Mechanism.

5.3.5 Interview Characteristics

In this section, we examine the interview characteristics for the PROMPT and NO PROMPT interviewers in Phase 02 to understand if the improved Prediction Mechanisms affected the interview as a whole in general. To do this, we analyze the interview duration and its variation among the PROMPT and NO PROMPT interview groups. We also examine the average number of activities per minute among the two interview groups to understand if there is any significant difference in the interview speed. Finally, we look at prediction-based analysis to report on any observable impact that the predictions had on the interview.

5.3.5.1 Interview Duration & Speed Analysis

First, we examine the interview duration and the number of activities recorded per minute for each of the interviewers and the groups in Phase 02. Table 49 lists the average interview duration, speed (as activities per minute) and average number of activities per interview statistics for the interviewers and the interview groups.

Id	Average interview time (minutes)	Average activities per minute	Average number of activities
I28 (PROMPT	12.600	2.017	23.80
I31 (PROMPT)	12.667	1.934	24.67
I29 (NO PROMPT)	17.000	1.869	25.90
I30 (NO PROMPT)	25.200	1.267	25.40
PROMPT	12.625	1.986	24.13
NO PROMPT	19.733	1.668	25.73

Table 49 Average interview time and average activities per minute statistics

From Table 49, we can observe that the PROMPT interviewers have a lower average interview time (12.625 mins versus 19.733) than NO PROMPT. This means that the PROMPT interviewers generally take less time than the NO PROMPT interviewers to complete the interviews. This is a desirable behavior for our instrument since interviews that take less time allow for faster completion. From Table 49, we also observe that the PROMPT interviewers have slightly better average activities per minute (1.986 versus 1.668) statistic than the NO PROMPT interviewers. This means that the PROMPT interviewers create more activities in the same time that it takes for the NO PROMPT interviewers. This is again a desirable effect as when the interviewers are able to create more activities faster, they are able to record faster and in turn make the interview more efficient. Thus, it would seem that the PROMPT interviewers were able to leverage the improvement in the instrument to conduct shorter and faster interviews – an important and desirable characteristic to conduct time diary surveys.

5.3.5.2 Predicted Precode Usage Analysis

Next, we analyze the predicted precodes usage to examine the usage of the predictions and to understand if the predictions being delivered through the precodes are useful or not. For this, Table 50 lists the average (avg.) precode statistics for the interviewers. In Table 50, the Avg. predictions made represents the average number of predictions made per interview for the corresponding interviewer(s). Similarly, the Avg. predictions clicked represents the average number of predictions clicked per interview and the Avg. precodes clicked represents the average number of precodes clicked per interview.

Id	Avg. predictions made	Avg. predictions clicked	Avg. precodes clicked
I28 (PROMPT)	106.000	12.500	63.200
I31 (PROMPT)	108.333	10.667	60.833
I29 (NO PROMPT)	114.300	10.90*	20.500
I30 (NO PROMPT)	118.600	6.60*	59.400
PROMPT	106.875	11.813	62.313
NO PROMPT	115.733	9.47	33.467

Table 50 Average predictions and precodes statistics for interviewr(s). * For NO PROMPT interviewers, the average prediction clicks represents the average number of times the interviewer selected the same precode as would have been predicted for the activity nam

From Table 50, we observe that there is predictions usage for the PROMPT interviewers (11.813 for PROMPT group). This means that the PROMPT interviewers clicked on the predictions made to enter in the activity name 11.8 times per interview on average. It must be noted here that there is a factor of 5 when considering the number of predictions made as up to 5 predictions may be made for an activity – thus with *n* activities in an interview, *5n* predictions can be made – however, only *n* predictions may be clicked (one for each of the *n* activities). This is a highly desirable observation as it means that the predictions made through the precodes are being successfully used by the interviewers to perform data entry. This observation when combined with the one wherein the PROMPT interviewers have shorter interviews (from Section 5.3.5.1) could potentially imply that the interviewers are able to use the precodes to perform data entry faster and in turn reduce the time taken to complete the interviews. This supports our design change decision to move the predictions to be delivered through the precodes. Further, this also supports of our framework's intention of delivering predictions to reduce data entry time and the interview time.

Another supporting observation from Table 50 is the higher average (almost double) precodes clicked for the PROMPT interviewers (62.313) when compared to the NO PROMPT interviewers (33.467). This means that the PROMPT interviewers preferred to use the precodes to enter in data almost twice the number of times as the NO PROMPT interviewers did. This higher precode usage by the PROMPT interviewers may be further attributed to the predictions being made on the precode itself and thus resulting in a higher number of activities created per minute. We also believe that the NO PROMPT interviewers' use of the precodes are being influenced by the absence of the predictions on the precodes as they use the precodes lesser (e.g., I29 has only 20.5 precodes clicked on average per interview as opposed to the average of 60 for the other interviewers). One possible reason for the NO PROMPT interviewers using the autocomplete more could be because they didn't have predictions on the precodes.

5.3.6 Post Interview Survey Analysis

The post interview survey was administered to the interviewer on successful completion of an interview. The survey is in a questionnaire format and consists of 8 questions, most of them with Likert scale type responses. Both the PROMPT and NO PROMPT interviews were followed by the same questionnaire. The questions in the post interview survey are listed in Appendix 7.3. In this section, we analyze the post interview survey question responses of the Phase 02 interviewers to understand if our observations regarding the predictions and the instrument's usefulness are reflected in the interviewer's feedback.

The questions that we are interested in to understand the effectiveness of the instrument with respect to the interviewers are question 1 (for PROMPT), question 4, and question 5. Each of these questions attempts to measure the impact of the instrument on the interview along different references based on the interviewers' opinion and personal evaluation of the interview. To help analyze the post interview survey data, we introduce a numerical value for each of the options for the question ranging from -2 to 2. This allows us to compute the average response value for a question aggregating the interviewers' responses. This average response value takes the score for each individual response and calculates their average to generate an average score. It must be noted here that the average score should not be used to make direct inferences about the interviewers' average response. This is because, the Likert scale is not an interval scale and thus the numerical values do not represent valid differences. For example, taking the average of the values for Strongly Agree (+2) and Strongly Disagree (-2) gives us 0 – which we may attempt to use to claim that the responses indicate neither agreement nor disagreement (0 = Neither Agree nor Disagree). However, this did not capture the fact that the responses were from two extremes of the scale. Nonetheless, this measure provides us with a way to understand where the agreement/disagreement tendency of the responses lies. Thus, for the example above, we could say that the particular average (0) shows that the responses do not lean to favor agreement or disagreement. Combining this with a frequency distribution would help us understand how to better interpret the average score and thus the interviewers' feedback of the instrument.

Question 1 denotes *the interviewers' opinion on whether the predictions made were useful to the interviewer in the interview*. The average score of the PROMPT interviewers for this question thus represents if the interviewers are generally leaning towards agreeing or disagreeing that the predictions were useful to them during the interview. This question is in context only for the treatment group interviewers (PROMPT) as it asks about the predictions, which only the PROMPT interviewers would receive. Table 51 details the distribution of the responses for question 1 for the PROMPT interviewers. Figure 45 displays the frequency distribution of the response items for question 1 of the post interview survey based on the PROMPT interviewers' responses.

Q. 1	**Strongly Disagree**	**Disagree**	**NAND***	**Agree**	**Strongly Agree**	**Average Score**
28 (P)	1	2	0	7	0	0.30
31 (P)	0	0	0	6	0	1.00
PROMPT	1	2	0	13	0	0.56

Table 51 Response distribution for question 1. *NAND – Neither Agree nor Disagree

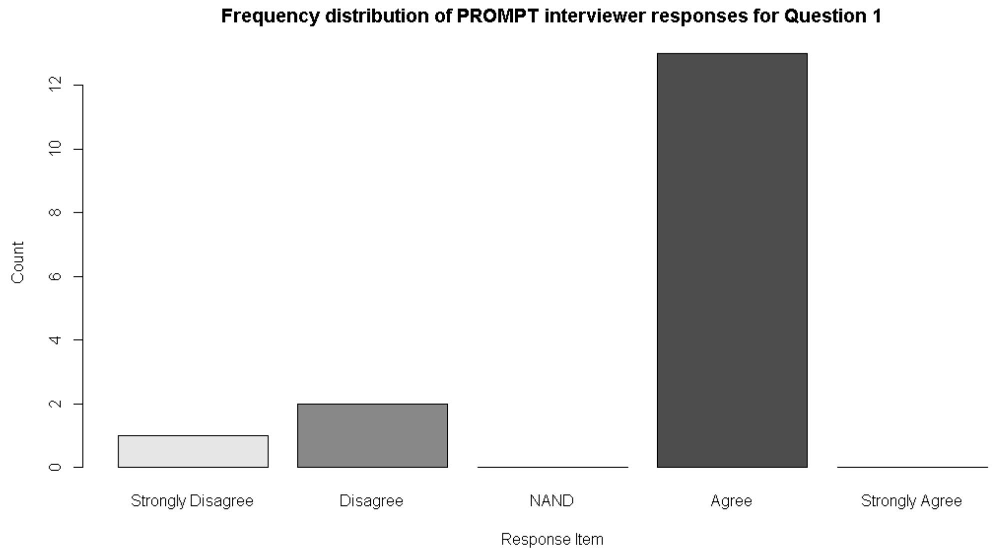

Figure 45 Frequency distribution of the responses of the PROMPT interviewers for question 1

From Table 51 and Figure 45, we observe that the most common response for question 1 is 'Agree' (13 of 16). The average score for the PROMPT interviewers is 0.56 which implies that they generally leaned more towards agreeing that the predictions were useful to them during the interview than disagreeing. This feedback from the PROMPT interviewers is in line with our observations in Section 5.3.4 and Section 5.3.5 based on the data that the predictions were enabling the PROMPT interviewers to conduct the interviews better. Combining the observations, we may imply that the changes performed on the instrument with respect to the predictions delivery were successfully able to induce a positive response and were useful to the interviewers.

Question 4 asks *the interviewers whether they believe that the instrument had a significant positive impact on the quality of the interview*. This represents the instrument's impact on the interview in a positive manner with respect to the interviewer. The positive impact could be in terms of efficiency, usefulness and/or effectiveness in assisting the interviewer based on the interviewer's interpretation of the question. Table 52 details the response distribution for question 4 for the Phase 02 interviewers.

Q. 4	Strongly Disagree	Disagree	NAND*	Agree	Strongly Agree	Average Score
28 (PROMPT)	0	0	4	6	0	0.60
31 (PROMPT)	0	0	1	5	0	0.83
29 (NO PROMPT)	1	5	4	0	0	-0.70
30 (NO PROMPT)	0	0	1	3	1	1.00
PROMPT	0	0	5	11	0	0.69
NO PROMPT	1	5	5	3	1	-0.13
ALL	1	5	10	14	1	0.29

Table 52 Response distribution for post interview survey question 4 for Phase 02 interviewers. *NAND Neither Agree nor Disagree

From Table 52, we observe that the majority response for question 4 is 'Agree' (14) and the average score for the interviewers is 0.29. Furthermore, PROMPT interviewers had the most 'Agree' responses (11) and the average score for the PROMPT interviewers is 0.69. The top NO PROMPT interviewer responses, however, are divided among 'Neither Agree nor Disagree' (5 instances) and 'Disagree' (5 instances) with an average score of -0.13.

Thus the PROMPT interviewers were leaning towards 'agreeing' that the instrument had a significant positive impact on the interviews since their average score of 0.69 is close to 1 (for 'Agree'). This ties in with our previous observations and findings that the predictions, which only the PROMPT interviewers received, were helpful in assisting the interviewers conduct the interviews better.

The NO PROMPT interviewers however, are leaning towards 'disagreeing' slightly, but were generally 'neither agreeing nor disagreeing' that the instrument had a significant positive impact on the interviews as their average score is less than 0 (which signifies 'Neither Agree nor Disagree'). However, we observe from Table 52 that of the NO

PROMPT interviewers, Interviewer 29 is the influencing interviewer with the most negative score (-0.70). While we are unable to explain with certainty why interviewer 29 believes that the instrument did not have a significant positive impact on the interview, we suspect that, in the interviewer's interpretation of the question, the instrument may not have assisted the interviewer significantly in the interview; as this interviewer is a NO PROMPT interviewer, this interpretation may be understandable.

Question 5 in the post interview survey asks *the interviewer whether the instrument had a significant negative impact on the quality of the interview*. This question serves to act as the complement of question 4 and can thus be used to verify if the response is within the complement. Furthermore, this question allows us to explicitly obtain feedback on whether the instrument effected the interview in a negative way. Table 53 details the response distribution for question 5 in the post interview survey.

Q. 5	Strongly Disagree	Disagree	NAND*	Agree	Strongly Agree	Average Score
28 (PROMPT)	0	7	3	0	0	-0.70
31 (PROMPT)	0	6	0	0	0	-1.00
29 (NO PROMPT)	0	1	8	0	1	0.10
30 (NO PROMPT)	1	3	0	1	0	-0.80
PROMPT	0	13	3	0	0	-0.81
NO PROMPT	1	4	8	1	1	-0.20
ALL	1	17	11	1	1	-0.52

Table 53 Response distribution for post interview survey question 5

From Table 53, we observe that the most favored response is 'Disagree' among all the interviewers, obtaining 17 of 31 responses and the average score for all the interviewers' responses is -0.52. The PROMPT interviewers had 'Disagree' for their top response (13) and an average score of -0.81. The NO PROMPT interviewers had 'Neither Agree nor Disagree' has their most favored response (8) and an average score of -0.20.

This is consistent with question 3's response and the observed better interviews conducted by the PROMPT interviewers as compared to the NO PROMPT interviewers. This allows us to say that the instrument, with its predictions delivered through the precodes, was able to provide assistance to the PROMPT interviewers in conducting better time diary surveys that also resulted in good quality data.

The lack of a strong disagreement or agreement from the NO PROMPT interviewers regarding the instrument's negative impact suggests that the NO PROMPT interviewers may have expected more from the instrument or faced external difficulties while conducting the interviews. The lack of a strong disagreement is evidenced by the NO PROMPT interviewers' average score for this question of -0.20. This means that the NO PROMPT interviewers did not find the instrument particularly negative in influence, but wasn't strong enough to be disagreed with. This may show a stronger reaction from the PROMPT interviewers as opposed to the NO PROMPT interviewers opposing a negative effect by the instrument which can be interpreted as that the instrument positively affected the PROMPT interviewers stronger than the NO PROMPT interviewers felt it was negative. Furthermore, if interviewer 29's responses for question 3 and question 4 are not considered, we observe that the other interviewers are more consistent in their responses to the two complementary questions. Table 54 lists the responses of the interviewers to question 4 and question 5 to illustrate this.

Interviewer	**Question**	**Responses**					
		Strongly Disagree	**Disagree**	**NAND**	**Agree**	**Strongly Agree**	**Average Score**
28 (PROMPT)	4 (Positive impact)	0	0	4	6	0	0.60

	5 (Negative impact)	0	7	3	0	0	-.70
31 (PROMPT)	4 (Positive impact)	0	0	1	5	0	0.83
	5 (Negative impact)	0	6	0	0	0	-1.00
29 (NO PROMPT)	4 (Positive impact)	1	5	4	0	0	-0.70
	5 (Negative impact)	0	1	8	0	1	0.10
30 (NO PROMPT)	4 (Positive impact)	0	0	1	3	1	1.00
	5 (Negative impact)	1	3	0	1	0	-0.80

Table 54 Response distribution of Phase 02 interviewers for question 4 and question 5 from the post interview survey shown together for comparison

From Table 54, we observe that interviewer 29 has a score of -0.70 for question 4 (positive impact by instrument) and a score of 0.10 for the question 5 (negative impact by instrument). This means that according to interviewer 29, the instrument did not introduce a significant positive impact on the interviews, which we suspect was based on how the interviewer interpreted the meaning of positive impact; but they also believe that the instrument was not significantly detrimental to the interview. This adds weight to our suspicion that interviewer 29 interpreted the positive impact in question 4 strongly.

5.3.7 Phase 02 Summary

From the analysis of the partially collected data from Phase 02 of the experiment, we were able to understand that the introduction of the design change wherein, the predictions are being delivered through the precodes, introduced positive effects on the PROMPT interviewers. We were first able to show that the instrument continued to perform well as a time diary survey instrument in Phase 02 with the design changes, using our data quality analysis in Section 5.3.3. With the interviewer analysis in Section

5.3.4, we were able to understand that the PROMPT interviewers were able to create activities faster. This provides us with encouraging evidence that the improved Prediction Mechanisms were able to reduce the data entry times for the interviewers, making the interviews faster. Section 5.3.5 further strengthened the positive effects of the Prediction Mechanisms by demonstrating that the PROMPT interviewers were able to complete interviews faster and perform data entry faster (and create more activities per minute) using the precodes. Finally, in Section 5.3.6, we analyzed the post interview survey responses submitted by the interviewers and were able to gather feedback that confirms the observations based on analyzing the response data and paradata that, the PROMPT interviewers felt that the instrument provided good assistance and introduced a positive impact on the interviews by improving the interview conduction.

5.3.8 Limitations

As mentioned in Section 5.3.2, Phase 02 did not reach completion as of yet. Of the 48 targeted interviews, only 31 interviews had been completed. Understandably, this could introduce issues with the data and the subsequent analyses made. Thus, we present the possible limitations of the analysis based on Phase 02's current data.

1. The lack of all the interview sessions could introduce imbalances in the data especially when compared with Phase 01.

2. The interviewers in Phase 02 are not the same interviewers who participated in Phase 01 – this could introduce interviewer specific effects.

3. The absence of the complete data for the interviewers in Phase 02 could be causing a lack of observable statistical significances between the interviewer groups.

4. Finally, Phase 02 was spread out for a longer duration than Phase 01, and this could have prevented the interviewers from gaining familiarity and experience working with the instrument and the process of conducting time diary surveys due to the lack of continuous involvement.

5.4 Conclusions

We implemented a prototype instrument based on our proposed framework to work in interviewer-assisted mode (IAM) and performed two phases of experimental studies to empirically understand how it can improve time diary surveys administration under a computer assisted telephone interview (CATI) setup. The objectives of Phase 01 were to:

1. Determine if the framework's instrument implementation performed well as a time diary survey instrument,

2. Study the effects of using the different implemented Interaction and Knowledge Engineering Mechanisms. These include:

    a. The Prediction Mechanisms, and

    b. Different Interaction Mechanisms for data entry

Based on the results obtained from Phase 01 and the subsequent analysis of the response data and paradata, we were able to demonstrate the positive qualities of our instrument and the overlaying framework in achieving the objectives of Phase 01. We were able to show that the data quality and the goodness of the instrument was comparable to ATUS, 2010 – a known good quality time diary survey instrument (Section 5.2.3). We then used proxies to overcome the lack of ground truth to understand how the Prediction Knowledge Engineering Mechanism makes accurate and timely predictions that could be of use to the interviewer (Section 5.2.4). We also examine how the delivery of the predictions by the Prediction Interaction Mechanism failed to achieve usefulness due to a design flaw, which we correct for in Phase 02. We also examined and analyzed the data further to confirm the effects that the predictions had on the PROMPT interviewers and were able to notice that, while the predictions did not play a primary influencing role, they were able to improve the performance of the PROMPT interviewers a little when compared to the performance of the NO PROMPT interviewers (Section 5.2.5).

Following a preliminary analysis of Phase 01 data, we improved the Prediction Interaction Mechanism and integrated it with the Precode Interaction Mechanism and incorporated a feedback survey for the interviewers known as the post interview survey before the start of Phase 02. The objectives of Phase 02 were:

1. To confirm that the instrument implementation continued to perform as good quality time diary survey instrument,

2. To understand the usage of the improved Prediction Interaction Mechanism by the PROMPT interviewers and examine difference in performances between the PROMPT and NO PROMPT interviewers, and

3. To obtain feedback from the interviewers regarding the instrument so to report on the instrument's performance based on the opinion of the users – the interviewers.

Phase 02 only achieved partial completion, with data available from 31 of the targeted 48 interviews. Based on the analysis of the available data, we were able to show that the instrument continued to perform well as a time diary survey instrument (Section 5.3.3). We were also able to demonstrate tentatively that the improved Prediction Interaction Mechanism was able to introduce an improvement in the performance of the PROMPT interviewers as compared to the NO PROMPT interviewers (Section 5.3.4). This performance improvement was in terms of being able to create activities faster – an indication of decreased cognitive load on the interviewer. We were also able to determine an overall improvement in the interview performance of the PROMPT interviewers with encouraging evidence that showed that they were able to complete interviews faster with reduced data entry times (Section 5.3.5). Finally, we were able to examine the post-interview survey responses and strengthen the observations made based on the analysis of the data (Section 5.3.6), that the instrument was able to assist the PROMPT interviewers well and provided a significant positive impact on the interviews.

Thus, with our implementation of the intelligent integrated multi-mode time diary survey framework in IAM mode, and our experimental studies of this implementation we can conclude that:

1. The instrument improves the interview process as intended and increases the data quality of the response data collected, when compared to a known time diary survey (ATUS, 2013). This provides us with evidence that the intelligent framework designed to assist the interviewer (in IAM) works as intended.

2. The framework's mechanisms contribute towards reducing the cognitive load on the interviewer and promotes faster data entry and reduced interview time. The Interaction Mechanisms for data entry – such as the Autocomplete Interaction Mechanism, the Precode Interaction Mechanism and the Timeline Interaction Mechanism provided the interviewers with multiple ways to enter data and was used by the interviewers to successfully record data during the interview based on their requirements.

3. The Prediction Interaction Mechanism and its improved version in Phase 02, provided assistance to the interviewers by allowing them to quickly identify and enter the activities by highlighting the predicted next activities distinctly in yellow.

4. The framework's implementation of IAM worked well and provided us with elicited knowledge of how the interviewers conducted the interviews.

5. The framework's flexibility and ease of modification was exemplified by the design change that was implemented in the way predictions were delivered between Phase 01 and Phase 02. The change, which took approximately 8 man

hours of work, and included rewriting the components and testing them involved only the following code changes listed in Table 55.

Mechanism	Phase 01	Phase 02	Affected files
Prediction Knowledge Engineering Mechanism	Predictions generated were based on time of day and previous activity.	Predictions generated were based on previous activity alone.	AgentBase.java
Prediction Interaction Mechanism	Predictions were rendered on a separate prompt panel.	Predictions were forwarded to the Precode Interaction Mechanism.	atus-prompt.js
Precode Interaction Mechanism	Displays the precodes in the precode panel	Displays the precodes in the precode panel and accepts the list of predictions from the Prediction Interaction Mechanism and applies a yellow highlight on their precode equivalents.	atus-internal-vms.js

Table 55 List of changes to the mechanisms and their corresponding implementation files for the design change in delivering predictions between Phase 01 and Phase 02

6. Thus finally, we conclude that the intelligent integrated multi-mode time diary survey framework was successfully implemented in its interviewer assisted mode and paves the way to the next step implementation of its self-administered mode.

Chapter 6: Conclusions

Our research is the design and development of an intelligent integrated framework that is suitable to administer time diary surveys (TDS) under two modes: an *interviewer assisted* (IA) mode and a *self-administered* (SA) mode. In the interviewer assisted mode, the system interacts with the interviewer who interacts with the respondent (directly or over the telephone). In the self-administered mode, the respondent directly interacts with the system. The objective of the framework thus brings about two primary questions– (1) how to model the interview process and (2) how to interact with the user within the rules of the survey domain.

The question of how to model the interview process is raised due to the nature of the problem that the framework is attempting to solve. TDS are essentially conversational surveys wherein either the interviewer or the respondent (or both) primarily control how the survey proceeds depending on the administration mode. The other aspect of the problem of how to interact with the user is more open-ended. The onus of keeping the respondent engaged usually rests with the interviewer who uses their expert interviewing knowledge to keep the interview on track as much as possible. Eliciting the required responses is the objective of the interviewer and he or she may employ conversational techniques and recall techniques to guide the respondent through the interview. These tasks shift on to the instrument in self-administered mode.

In our work, we have described a framework that can assist the user in completing time diary surveys and that can be adapted to work in both interviewer-assisted mode (IAM) and self-administered mode (SAM). For this, we have proposed our intelligent integrated multi-mode time diary survey framework, that would use two sets of overreaching components called Mechanisms – Interaction Mechanisms (IxM) and Knowledge Engineering Mechanisms (KEM). The Interaction Mechanisms essentially deal with the problems of interacting with the different types of users while the Knowledge Engineering Mechanisms focus on modeling the interview process and the users. Thus, the mechanisms have been designed to work synergistically to solve their respective problems and then interact with each other to create a working implementation that solves the problem as a whole. This separation, while in no way complete, allows for division of the problems in such a manner that it reduces each mechanism's individual problem to singular units. For example, in our proposed framework, the Prediction Mechanisms consists of two component mechanisms – the Prediction Knowledge Engineering Mechanism and the Prediction Interaction Mechanism. The Prediction Knowledge Engineering Mechanism will thus be allowed to solve the problem of what predictions to make based on the data available and generates a list of predictions to be made. The Prediction Interaction Mechanism then leverages the predictions generated by the Prediction Knowledge Engineering Mechanism and delivers them to the user in an efficient manner – and thus deals with the problem of how to interact with the user. This modularization allows the framework to be adaptable, extendible and scalable.

6.1 Contributions

The primary contribution of our work is in paving the way for the employment of Computer Science in the niche fields of Intelligent User Interfaces (IUI) and Recommender Systems (RS) with respect to restrictive environments. Our contribution to these two fields are unique in that, we are dealing with a unique domain that is characterized by limitations imposed by biases, restricted feedback and that involves in knowledge elicitation from participants of different motivations. Time diary survey is the domain that we deal with in our work – but the concepts can be extended to domains with similar restrictions. Furthermore, while Computer Science technologies have been used in time diary surveys, they have mostly been approached from the point of view of surveys. In our work, we approach the surveys from the Computer Science point of view – thus we are gearing towards providing a solution to work in such a restrictive domain.

Our next contribution is the integrated framework where we combine two different kinds of systems by distributing tasks between different mechanisms. This allows the framework to switch mechanisms and handle distinct and different users within a core framework structure. This paves the way for future research work that can extend and add mechanisms to deal with new problems or adapt existing mechanisms to deal with similar problems. The integrated framework is geared towards handling two kinds of users in two modes, IAM and SAM and provides ways to handle both modes by sharing common problems while handling distinctly different problems with specific mechanisms. Mechanisms that only deal with one mode can be simply turned off without requiring extensive rework.

Our framework's prototype instrument is a practical contribution that illustrates how the framework can be implemented with the mechanisms that work in interviewer assisted mode in a computer assisted telephone interview (CATI) setup. This implementation also deals with assisting the interviewers (without biasing them), enabling visualization of data in real time that the interviewers can use to assist the respondents, providing status updates to the interviewers to allow them to keep track of the interview progress and providing multiple ways to enter data based on the interviewer's preference. All these contribute towards reducing the interviewer's cognitive load while conducting interviews so that they can engage more with the respondents.

We also contribute towards the future work in this domain with our experimental results and our analysis of the collected response data and paradata. One of the end products of our experimental studies is the transcripts of the interviews conducted using instrument in IAM which contributes towards building the SAM implementation by providing elicited expert knowledge regarding the interviews. These transcripts are currently being used to research on how the interviewers conduct interviews and how the system can leverage this knowledge.

Our contribution to the field of survey research and methodology, especially time diary surveys is in the areas of CATI and adaptive designs for surveys. While conventional CATI systems tend to be focused on data entry, our instrument prototype adds an intelligent component to it that can actually *assist* the interviewer by reducing their cognitive load during the interview and thus improving the interview in terms of speed, time required and data quality. We also contribute towards paradata tracking and analysis of time diary survey interviews. This is in terms of gathering and storing of paradata, which is data about how the response data was collected, and the subsequent analysis of this paradata that can be leveraged to further improve the time diary survey interview process. We also contribute towards understanding how to use historical survey response data to deal with cold-start problems by reporting on how such data can be converted to domain knowledge for the system.

We also contribute towards understanding interviewer modeling and respondent modeling and how the two apparently distinct users (interviewers and respondents) can be viewed as one type of user with distinct characteristics. These distinct characteristics are the input problems for different mechanism in our framework and thus creates an intelligent, integrated multi-mode framework for time diary surveys. Our implementation of this framework in IAM sets the starting steps to integrate multiple modes of surveys so that response data can be shared across systems to improve them. This is a unique contribution as, currently, CATI systems and self-administered survey systems work independently and produce non-compatible response data which must be integrated through an offline process known as homogenization.

Finally, we also contribute towards the combined domain of survey informatics by implementing a framework that enables the use of tracked paradata from interviews to improve how the system interacts with the users.

6.2 Future Work

First, we need a more detailed set of additional tests that can be used to identify the effects of each Interaction Mechanism more closely. In our work, we were able to examine the effects of the Prediction Interaction Mechanism and the mechanisms used for data entry—however, the unavailability of more interviews reduced the data set size within a phase. Performing experiments with more interviewers and respondents to specifically test the effectiveness of the mechanisms would enable a better understanding of how the mechanisms affect the interviewer—and thus help improve the instrument more. For testing the Knowledge Engineering Mechanisms (KEM), a larger data set of collected results would help evaluate the differences brought by each KEM by taking separate control and treatment groups.

A more influential future work would be extending the framework into its next potential stage, where the time diary survey can be administered directly to the respondent – known as the self-administration mode (SAM). With this, the respondent directly interacts with the instrument to record their own activities. Without the guidance of the interviewer in SAM, the instrument must aim to guide the respondent through the survey – providing prompts and probes where required, while the respondent completes the survey. As mentioned in Chapter 1, the respondents are not as motivated to use the instrument as the interviewers are, and this provides the setting for the challenge of working directly with the respondents. This future work piece can potentially implement many mechanisms that are geared towards the respondents and/or modify the behavior of existing mechanisms to work with the respondents. As an example, the Precode IxM in our implementation may *not* be presented directly to the respondent as it is a potential source of satisficing. Research is needed to accurately identify if the Precode IxM should exist as such (and risk satisficing) or if it (Precode IxM) must be modified (for example, display the list only when the respondent has not entered data for an amount of time, etc.). Furthermore, newer IxMs and KEMs can be implemented catering to respondents based on how they interact with interviewers – for example, the respondent's speech can potentially be converted to text and natural language processing can be applied to *understand* what the respondent wants or wishes to record. These mechanisms may chain themselves to other mechanisms – from the example, the NLP based processing can be chained to predictive or corrective mechanisms that can work the way interviewers do by correcting mistakes and/or probing for more information.

While the overall SAM implementation might seem cumbersome, our framework provides a way to examine how the components need to interact with one another and what is to be expected from the mechanisms; thus making the development of the instrument's SAM relatively simpler. The integration of SAM and IAM can be achieved using simple in-instrument switches when interviews are created. During the initialization of the instrument, the mechanisms can turn on or off depending on the mode switch. For example, the Precode IxM may be turned on when the user is an interviewer, but can be turned off (or even run as a modified mechanism) when the user is a respondent. This would thus enable the eventual building of an instrument that would perform in both modes in an integrated manner – the modes being interviewer-assisted mode (IAM) and self-administered mode (SAM).

Another potential future work would be in adding the virtual interviewer from Conrad, 2015's work to our framework in self-administered mode. Their work used a "wizarded" virtual interviewer, that was controlled by a hidden researcher. Our framework provides a way to add a virtual interviewer as an Interaction Mechanism and then have supporting Knowledge Engineering Mechanisms to provide the assistance to the respondent through the virtual interviewer. This would offer the virtual interviewer with supporting intelligent components so that predictions, probes, and other assistive features can be delivered to the respondent in a more 'human interviewer' like manner. In a time diary survey, just keeping the respondent engaged and moving forward with the interview would be a significant victory in the development of self-administered time diary surveys.

Looking further ahead, since the framework essentially builds on top of a web-based communication system, it can be ported and deployed with minimal changes on smartphones and other screen based devices such as tablets. Since the design (the positioning and sizing) of the instrument is the primary change, the framework can work relatively without much modifications of the mechanisms underneath the corresponding implementation. One of the major issues with porting from large screen interfaces to mobile-based small screen interfaces is the requirement for major re-designing to fit the smaller screens. The framework provides a way to handle this because of the way Interaction Mechanisms work; wherein, a set of mobile-based Interaction Mechanisms can be created for small screen interfaces that are modified versions of the normal Interaction Mechanisms. Then, the corresponding Interaction Mechanisms can be switched based on the screen size to adjust accordingly. While this may seem counter-intuitive (due to the creation of more Interaction Mechanisms), it must be realized that with smaller screens, certain Interaction Mechanisms, such as the Timeline Interaction Mechanism may not be able to function at all without a wider screen to display the entire 24-hour duration. However, by adding in Interaction Mechanisms specifically for the smaller screens, we can employ the use of the Knowledge Engineering Mechanisms, which would remain unchanged, and adapt accordingly. This would at least reduce the time and effort required to port the instrument to smaller mobile screens.

Thus, this outlines the path ahead for the instrument as it attempts to integrate the two modes together and provide intelligent assistance to the user.

Chapter 7: Appendix

7.1 ATUS defined activities

6-digit activity code	Activity
10101	Sleeping
10102	Sleeplessness
10199	Sleeping, n.e.c.*
10201	Washing, dressing and grooming oneself
10299	Grooming, n.e.c.*
10301	Health-related self care
10399	Self care, n.e.c.*
10401	Personal/Private activities
10499	Personal activities, n.e.c.*
10501	Personal emergencies
10599	Personal care emergencies, n.e.c.*
19999	Personal Care, n.e.c.*
20101	Interior cleaning
20102	Laundry
20103	Sewing, repairing, & maintaining textiles
20104	Storing interior hh items, inc. food
20199	Housework, n.e.c.*
20201	Food and drink preparation
20202	Food presentation
20203	Kitchen and food clean-up
20299	Food & drink prep, presentation, & clean-up, n.e.c.*
20301	Interior arrangement, decoration, & repairs
20302	Building and repairing furniture
20303	Heating and cooling
20399	Interior maintenance, repair, & decoration, n.e.c.*
20401	Exterior cleaning
20402	Exterior repair, improvements, & decoration
20499	Exterior maintenance, repair & decoration, n.e.c.*
20501	Lawn, garden, and houseplant care
20502	Ponds, pools, and hot tubs
20599	Lawn and garden, n.e.c.*
20601	Care for animals and pets (not veterinary care)
20602	Walking / exercising / playing with animals
20699	Pet and animal care, n.e.c.*

20701	Vehicle repair and maintenance (by self)
20799	Vehicles, n.e.c.*
20801	Appliance, tool, and toy set-up, repair, & maintenance (by self)
20899	Appliances and tools, n.e.c.*
20901	Financial management
20902	Household & personal organization and planning
20903	HH & personal mail & messages (except e-mail)
20904	HH & personal e-mail and messages
20905	Home security
20999	Household management, n.e.c.*
29999	Household activities, n.e.c.*
30101	Physical care for hh children
30102	Reading to/with hh children
30103	Playing with hh children, not sports
30104	Arts and crafts with hh children
30105	Playing sports with hh children
30106	Talking with/listening to hh children
30108	Organization & planning for hh children
30109	Looking after hh children (as a primary activity)
30110	Attending hh children's events
30111	Waiting for/with hh children
30112	Picking up/dropping off hh children
30199	Caring for & helping hh children, n.e.c.*
30201	Homework (hh children)
30202	Meetings and school conferences (hh children)
30203	Home schooling of hh children
30204	Waiting associated with hh children's education
30299	Activities related to hh child's education, n.e.c.*
30301	Providing medical care to hh children
30302	Obtaining medical care for hh children
30303	Waiting associated with hh children's health
30399	Activities related to hh child's health, n.e.c.*
30401	Physical care for hh adults
30402	Looking after hh adult (as a primary activity)
30403	Providing medical care to hh adult
30404	Obtaining medical and care services for hh adult
30405	Waiting associated with caring for household adults
30499	Caring for household adults, n.e.c.*
30501	Helping hh adults
30502	Organization & planning for hh adults
30503	Picking up/dropping off hh adult

30504	Waiting associated with helping hh adults
30599	Helping household adults, n.e.c.*
39999	Caring for & helping hh members, n.e.c.*
40101	Physical care for nonhh children
40102	Reading to/with nonhh children
40103	Playing with nonhh children, not sports
40104	Arts and crafts with nonhh children
40105	Playing sports with nonhh children
40106	Talking with/listening to nonhh children
40108	Organization & planning for nonhh children
40109	Looking after nonhh children (as primary activity)
40110	Attending nonhh children's events
40111	Waiting for/with nonhh children
40112	Dropping off/picking up nonhh children
40199	Caring for and helping nonhh children, n.e.c.*
40201	Homework (nonhh children)
40202	Meetings and school conferences (nonhh children)
40203	Home schooling of nonhh children
40204	Waiting associated with nonhh children's education
40299	Activities related to nonhh child's educ., n.e.c.*
40301	Providing medical care to nonhh children
40302	Obtaining medical care for nonhh children
40303	Waiting associated with nonhh children's health
40399	Activities related to nonhh child's health, n.e.c.*
40401	Physical care for nonhh adults
40402	Looking after nonhh adult (as a primary activity)
40403	Providing medical care to nonhh adult
40404	Obtaining medical and care services for nonhh adult
40405	Waiting associated with caring for nonhh adults
40499	Caring for nonhh adults, n.e.c.*
40501	Housework, cooking, & shopping assistance for nonhh adults
40502	House & lawn maintenance & repair assistance for nonhh adults
40503	Animal & pet care assistance for nonhh adults
40504	Vehicle & appliance maintenance/repair assistance for nonhh adults
40505	Financial management assistance for nonhh adults
40506	Household management & paperwork assistance for nonhh adults
40507	Picking up/dropping off nonhh adult
40508	Waiting associated with helping nonhh adults
40599	Helping nonhh adults, n.e.c.*

49999	Caring for & helping nonhh members, n.e.c.*
50101	Work, main job
50102	Work, other job(s)
50103	Security procedures related to work
50104	Waiting associated with working
50199	Working, n.e.c.*
50201	Socializing, relaxing, and leisure as part of job
50202	Eating and drinking as part of job
50203	Sports and exercise as part of job
50204	Security procedures as part of job
50205	Waiting associated with work-related activities
50299	Work-related activities, n.e.c.*
50301	Income-generating hobbies, crafts, and food
50302	Income-generating performances
50303	Income-generating services
50304	Income-generating rental property activities
50305	Waiting associated with other income-generating activities
50399	Other income-generating activities, n.e.c.*
50401	Job search activities
50403	Job interviewing
50404	Waiting associated with job search or interview
50405	Security procedures rel. to job search/interviewing
50499	Job search and Interviewing, n.e.c.*
59999	Work and work-related activities, n.e.c.*
60101	Taking class for degree, certification, or licensure
60102	Taking class for personal interest
60103	Waiting associated with taking classes
60104	Security procedures rel. to taking classes
60199	Taking class, n.e.c.*
60201	Extracurricular club activities
60202	Extracurricular music & performance activities
60203	Extracurricular student government activities
60204	Waiting associated with extracurricular activities
60299	Education-related extracurricular activities, n.e.c.*
60301	Research/homework for class for degree, certification, or licensure
60302	Research/homework for class for pers. interest
60303	Waiting associated with research/homework
60399	Research/homework n.e.c.*
60401	Administrative activities: class for degree, certification, or licensure
60402	Administrative activities: class for personal interest

60403	Waiting associated w/admin. activities (education)
60499	Administrative for education, n.e.c.*
69999	Education, n.e.c.*
70101	Grocery shopping
70102	Purchasing gas
70103	Purchasing food (not groceries)
70104	Shopping, except groceries, food and gas
70105	Waiting associated with shopping
70199	Shopping, n.e.c.*
70201	Comparison shopping
70299	Researching purchases, n.e.c.*
70301	Security procedures rel. to consumer purchases
70399	Security procedures rel. to consumer purchases, n.e.c.*
79999	Consumer purchases, n.e.c.*
80101	Using paid childcare services
80102	Waiting associated w/purchasing childcare svcs
80199	Using paid childcare services, n.e.c.*
80201	Banking
80202	Using other financial services
80203	Waiting associated w/banking/financial services
80299	Using financial services and banking, n.e.c.*
80301	Using legal services
80302	Waiting associated with legal services
80399	Using legal services, n.e.c.*
80401	Using health and care services outside the home
80402	Using in-home health and care services
80403	Waiting associated with medical services
80499	Using medical services, n.e.c.*
80501	Using personal care services
80502	Waiting associated w/personal care services
80599	Using personal care services, n.e.c.*
80601	Activities rel. to purchasing/selling real estate
80602	Waiting associated w/purchasing/selling real estate
80699	Using real estate services, n.e.c.*
80701	Using veterinary services
80702	Waiting associated with veterinary services
80799	Using veterinary services, n.e.c.*
80801	Security procedures rel. to professional/personal svcs.
80899	Security procedures rel. to professional/personal svcs n.e.c.*
89999	Professional and personal services, n.e.c.*
90101	Using interior cleaning services

90102	Using meal preparation services
90103	Using clothing repair and cleaning services
90104	Waiting associated with using household services
90199	Using household services, n.e.c.*
90201	Using home maint/repair/décor/construction svcs
90202	Waiting associated w/ home main/repair/décor/constr
90299	Using home maint/repair/décor/constr services, n.e.c.*
90301	Using pet services
90302	Waiting associated with pet services
90399	Using pet services, n.e.c.*
90401	Using lawn and garden services
90402	Waiting associated with using lawn & garden services
90499	Using lawn and garden services, n.e.c.*
90501	Using vehicle maintenance or repair services
90502	Waiting associated with vehicle main. or repair svcs
90599	Using vehicle maint. & repair svcs, n.e.c.*
99999	Using household services, n.e.c.*
100101	Using police and fire services
100102	Using social services
100103	Obtaining licenses & paying fines, fees, taxes
100199	Using government services, n.e.c.*
100201	Civic obligations & participation
100299	Civic obligations & participation, n.e.c.*
100304	Waiting associated with using government services
100305	Waiting associated with civic obligations & participation
100399	Waiting assoc. w/govt svcs or civic obligations, n.e.c.*
100401	Security procedures rel. to govt svcs/civic obligations
100499	Security procedures rel. to govt svcs/civic obligations, n.e.c.*
109999	Government services, n.e.c.*
110101	Eating and drinking
110199	Eating and drinking, n.e.c.*
110201	Waiting associated w/eating & drinking
110299	Waiting associated with eating & drinking, n.e.c.*
119999	Eating and drinking, n.e.c.*
120101	Socializing and communicating with others
120199	Socializing and communicating, n.e.c.*
120201	Attending or hosting parties/receptions/ceremonies
120202	Attending meetings for personal interest (not volunteering)
120299	Attending/hosting social events, n.e.c.*
120301	Relaxing, thinking
120302	Tobacco and drug use

120303	Television and movies (not religious)
120304	Television (religious)
120305	Listening to the radio
120306	Listening to/playing music (not radio)
120307	Playing games
120308	Computer use for leisure (exc. Games)
120309	Arts and crafts as a hobby
120310	Collecting as a hobby
120311	Hobbies, except arts & crafts and collecting
120312	Reading for personal interest
120313	Writing for personal interest
120399	Relaxing and leisure, n.e.c.*
120401	Attending performing arts
120402	Attending museums
120403	Attending movies/film
120404	Attending gambling establishments
120405	Security procedures rel. to arts & entertainment
120499	Arts and entertainment, n.e.c.*
120501	Waiting assoc. w/socializing & communicating
120502	Waiting assoc. w/attending/hosting social events
120503	Waiting associated with relaxing/leisure
120504	Waiting associated with arts & entertainment
120599	Waiting associated with socializing, n.e.c.*
129999	Socializing, relaxing, and leisure, n.e.c.*
130101	Doing aerobics
130102	Playing baseball
130103	Playing basketball
130104	Biking
130105	Playing billiards
130106	Boating
130107	Bowling
130108	Climbing, spelunking, caving
130109	Dancing
130110	Participating in equestrian sports
130111	Fencing
130112	Fishing
130113	Playing football
130114	Golfing
130115	Doing gymnastics
130116	Hiking
130117	Playing hockey

130118	Hunting
130119	Participating in martial arts
130120	Playing racquet sports
130121	Participating in rodeo competitions
130122	Rollerblading
130123	Playing rugby
130124	Running
130125	Skiing, ice skating, snowboarding
130126	Playing soccer
130127	Softball
130128	Using cardiovascular equipment
130129	Vehicle touring/racing
130130	Playing volleyball
130131	Walking
130132	Participating in water sports
130133	Weightlifting/strength training
130134	Working out, unspecified
130135	Wrestling
130136	Doing yoga
130199	Playing sports n.e.c.*
130201	Watching aerobics
130202	Watching baseball
130203	Watching basketball
130204	Watching biking
130205	Watching billiards
130206	Watching boating
130207	Watching bowling
130208	Watching climbing, spelunking, caving
130209	Watching dancing
130210	Watching equestrian sports
130211	Watching fencing
130212	Watching fishing
130213	Watching football
130214	Watching golfing
130215	Watching gymnastics
130216	Watching hockey
130217	Watching martial arts
130218	Watching racquet sports
130219	Watching rodeo competitions
130220	Watching rollerblading
130221	Watching rugby

130222	Watching running
130223	Watching skiing, ice skating, snowboarding
130224	Watching soccer
130225	Watching softball
130226	Watching vehicle touring/racing
130227	Watching volleyball
130228	Watching walking
130229	Watching water sports
130230	Watching weightlifting/strength training
130231	Watching people working out, unspecified
130232	Watching wrestling
130299	Attending sporting events, n.e.c.*
130301	Waiting related to playing sports or exercising
130302	Waiting related to attending sporting events
130399	Waiting associated with sports, exercise, & recreation, n.e.c.*
130401	Security related to playing sports or exercising
130402	Security related to attending sporting events
130499	Security related to sports, exercise, & recreation, n.e.c.*
139999	Sports, exercise, & recreation, n.e.c.*
140101	Attending religious services
140102	Participation in religious practices
140103	Waiting associated w/religious & spiritual activities
140104	Security procedures rel. to religious & spiritual activities
140105	Religious education activities
149999	Religious and spiritual activities, n.e.c.*
150101	Computer use
150102	Organizing and preparing
150103	Reading
150104	Telephone calls (except hotline counseling)
150105	Writing
150106	Fundraising
150199	Administrative & support activities, n.e.c.*
150201	Food preparation, presentation, clean-up
150202	Collecting & delivering clothing & other goods
150203	Providing care
150204	Teaching, leading, counseling, mentoring
150299	Social service & care activities, n.e.c.*
150301	Building houses, wildlife sites, & other structures
150302	Indoor & outdoor maintenance, repair, & clean-up
150399	Indoor & outdoor maintenance, building & clean-up activities, n.e.c.*
150401	Performing

150402	Serving at volunteer events & cultural activities
150499	Participating in performance & cultural activities, n.e.c.*
150501	Attending meetings, conferences, & training
150599	Attending meetings, conferences, & training, n.e.c.*
150601	Public health activities
150602	Public safety activities
150699	Public health & safety activities, n.e.c.*
150701	Waiting associated with volunteer activities
150799	Waiting associated with volunteer activities, n.e.c.*
150801	Security procedures related to volunteer activities
150899	Security procedures related to volunteer activities, n.e.c.*
159999	Volunteer activities, n.e.c.*
160101	Telephone calls to/from family members
160102	Telephone calls to/from friends, neighbors, or acquaintances
160103	Telephone calls to/from education services providers
160104	Telephone calls to/from salespeople
160105	Telephone calls to/from professional or personal care svcs providers
160106	Telephone calls to/from household services providers
160107	Telephone calls to/from paid child or adult care providers
160108	Telephone calls to/from government officials
160199	Telephone calls (to or from), n.e.c.*
160201	Waiting associated with telephone calls
160299	Waiting associated with telephone calls, n.e.c.*
169999	Telephone calls, n.e.c.*
180101	Travel related to personal care
180199	Travel related to personal care, n.e.c.*
180201	Travel related to housework
180202	Travel related to food & drink prep., clean-up, & presentation
180203	Travel related to interior maintenance, repair, & decoration
180204	Travel related to exterior maintenance, repair, & decoration
180205	Travel related to lawn, garden, and houseplant care
180206	Travel related to care for animals and pets (not vet care)
180207	Travel related to vehicle care & maintenance (by self)
180208	Travel related to appliance, tool, and toy set-up, repair, & maintenance (by self)
180209	Travel related to household management
180299	Travel related to household activities, n.e.c.*
180301	Travel related to caring for & helping hh children
180302	Travel related to hh children's education
180303	Travel related to hh children's health
180304	Travel related to caring for hh adults

180305	Travel related to helping hh adults
180399	Travel rel. to caring for & helping hh members, n.e.c.*
180401	Travel related to caring for and helping nonhh children
180402	Travel related to nonhh children's education
180403	Travel related to nonhh children's health
180404	Travel related to caring for nonhh adults
180405	Travel related to helping nonhh adults
180499	Travel rel. to caring for & helping nonhh members, n.e.c.*
180501	Travel related to working
180502	Travel related to work-related activities
180503	Travel related to income-generating activities
180504	Travel related to job search & interviewing
180599	Travel related to work, n.e.c.*
180601	Travel related to taking class
180602	Travel related to extracurricular activities (ex. Sports)
180603	Travel related to research/homework
180604	Travel related to registration/administrative activities
180699	Travel related to education, n.e.c.*
180701	Travel related to grocery shopping
180702	Travel related to purchasing gas
180703	Travel related to purchasing food (not groceries)
180704	Travel related to shopping, ex groceries, food, and gas
180799	Travel related to consumer purchases, n.e.c.*
180801	Travel related to using childcare services
180802	Travel related to using financial services and banking
180803	Travel related to using legal services
180804	Travel related to using medical services
180805	Travel related to using personal care services
180806	Travel related to using real estate services
180807	Travel related to using veterinary services
180899	Travel rel. to using prof. & personal care services, n.e.c.*
180901	Travel related to using household services
180902	Travel related to using home main./repair/décor./construction svcs
180903	Travel related to using pet services (not vet)
180904	Travel related to using lawn and garden services
180905	Travel related to using vehicle maintenance & repair services
180999	Travel related to using household services, n.e.c.*
181001	Travel related to using government services
181002	Travel related to civic obligations & participation
181099	Travel rel. to govt svcs & civic obligations, n.e.c.*
181101	Travel related to eating and drinking

181199	Travel related to eating and drinking, n.e.c.*
181201	Travel related to socializing and communicating
181202	Travel related to attending or hosting social events
181203	Travel related to relaxing and leisure
181204	Travel related to arts and entertainment
181205	Travel as a form of entertainment
181299	Travel rel. to socializing, relaxing, & leisure, n.e.c.*
181301	Travel related to participating in sports/exercise/recreation
181302	Travel related to attending sporting/recreational events
181399	Travel related to sports, exercise, & recreation, n.e.c.*
181401	Travel related to religious/spiritual practices
181499	Travel rel. to religious/spiritual activities, n.e.c.*
181501	Travel related to volunteering
181599	Travel related to volunteer activities, n.e.c.*
181601	Travel related to phone calls
181699	Travel rel. to phone calls, n.e.c.*
181801	Security procedures related to traveling
181899	Security procedures related to traveling, n.e.c.*
189999	Traveling, n.e.c.*
500101	Insufficient detail in verbatim
500103	Missing travel or destination
500104	Recorded simultaneous activities incorrectly
500105	Respondent refused to provide information/"none of your business"
500106	Gap/can't remember
500107	Unable to code activity at 1st tier
509999	Data codes, n.e.c.*

Table 56 ATUS (2010) activities list. n.e.c* - not elsewhere classified

7.2 ATUS Probing Charts

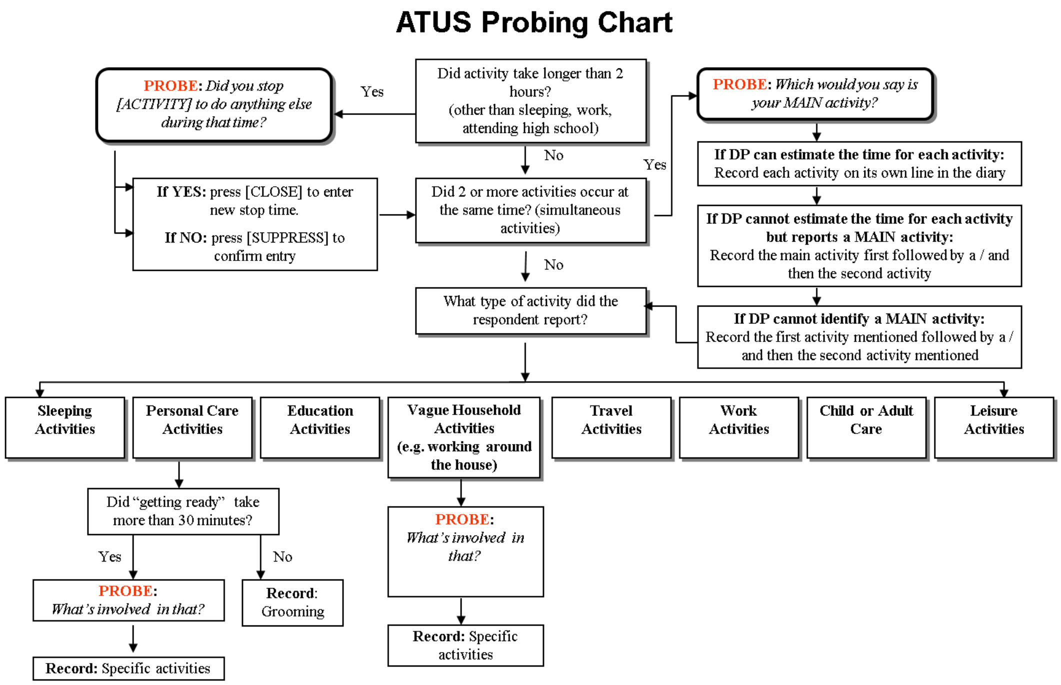

Figure 46 ATUS general probing rules

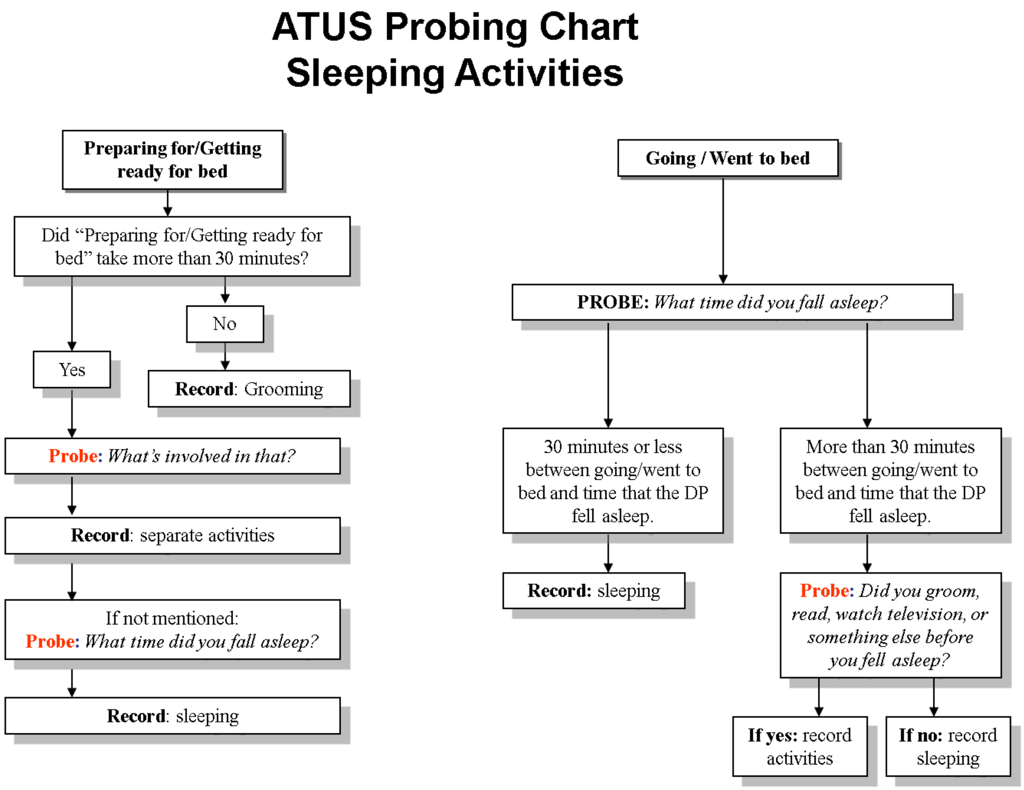

Figure 47 ATUS sleeping activities probes

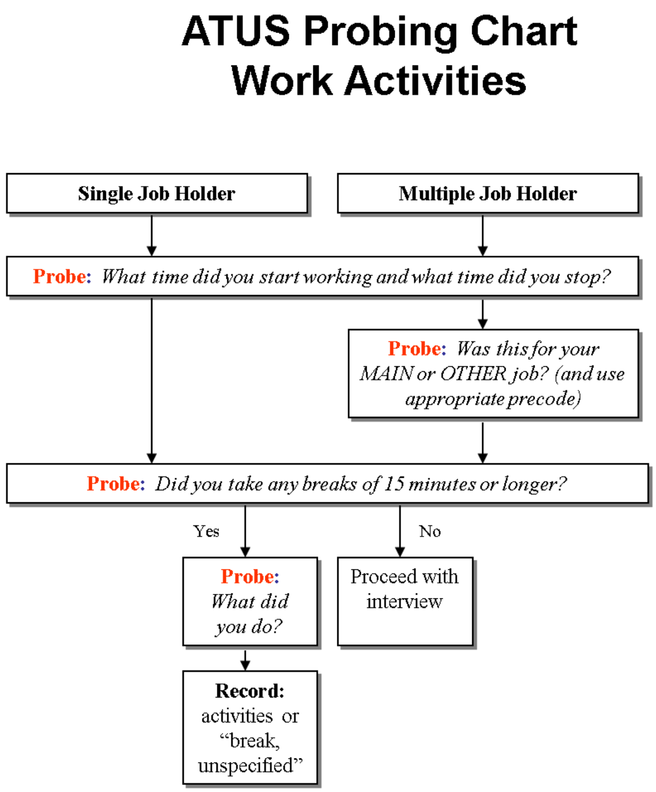

Figure 48 ATUS work activities probing chart

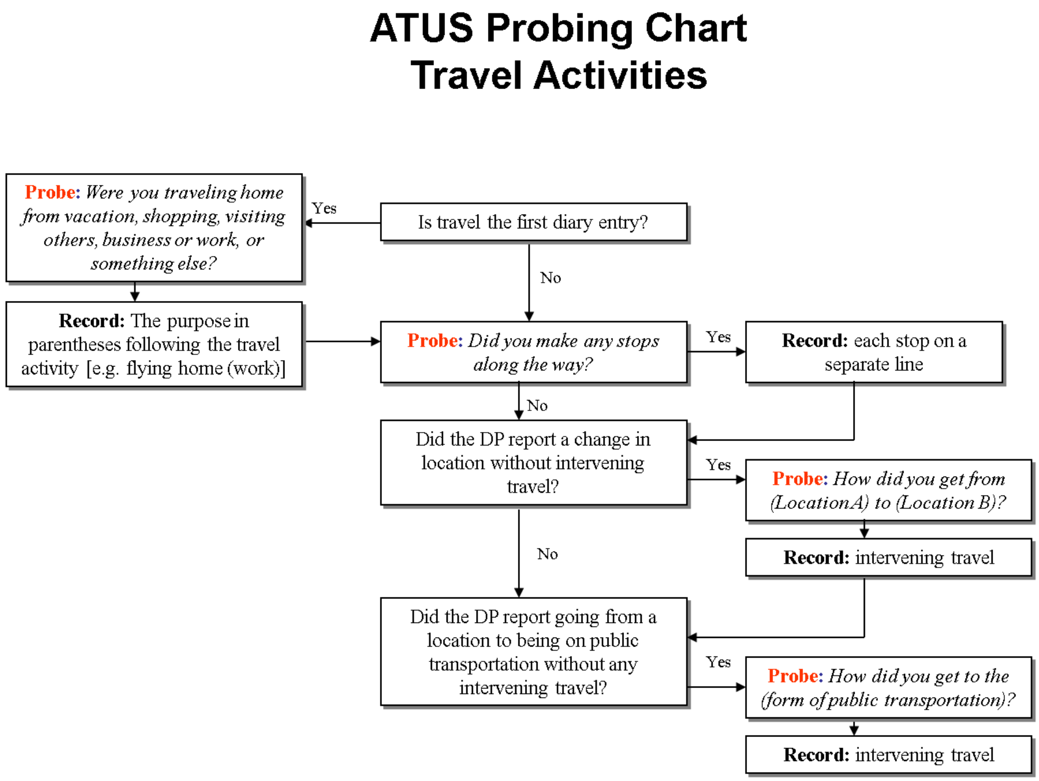

Figure 49 ATUS traveling activity probe chart

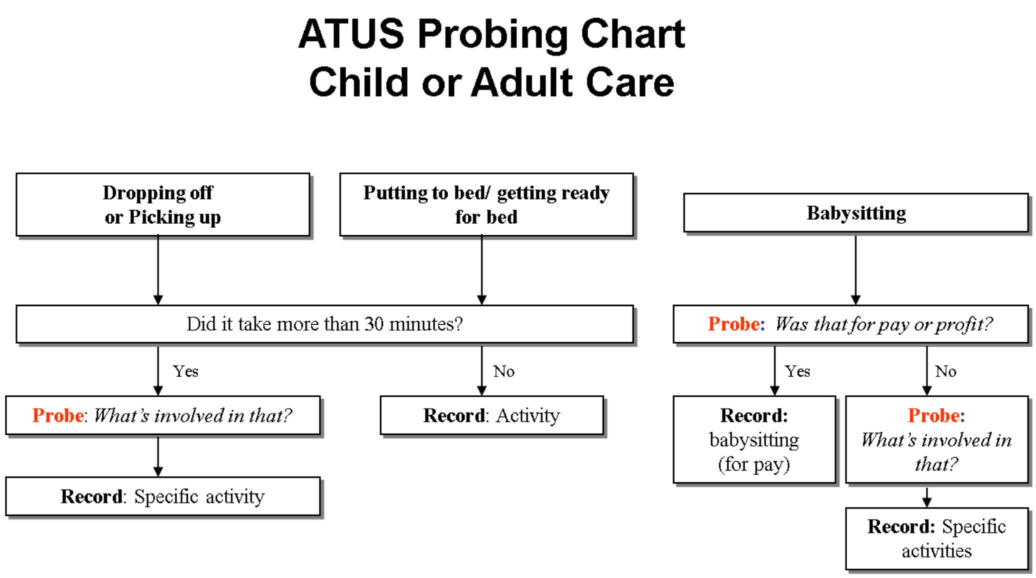

Figure 50 ATUS child or adult care activities probe chart

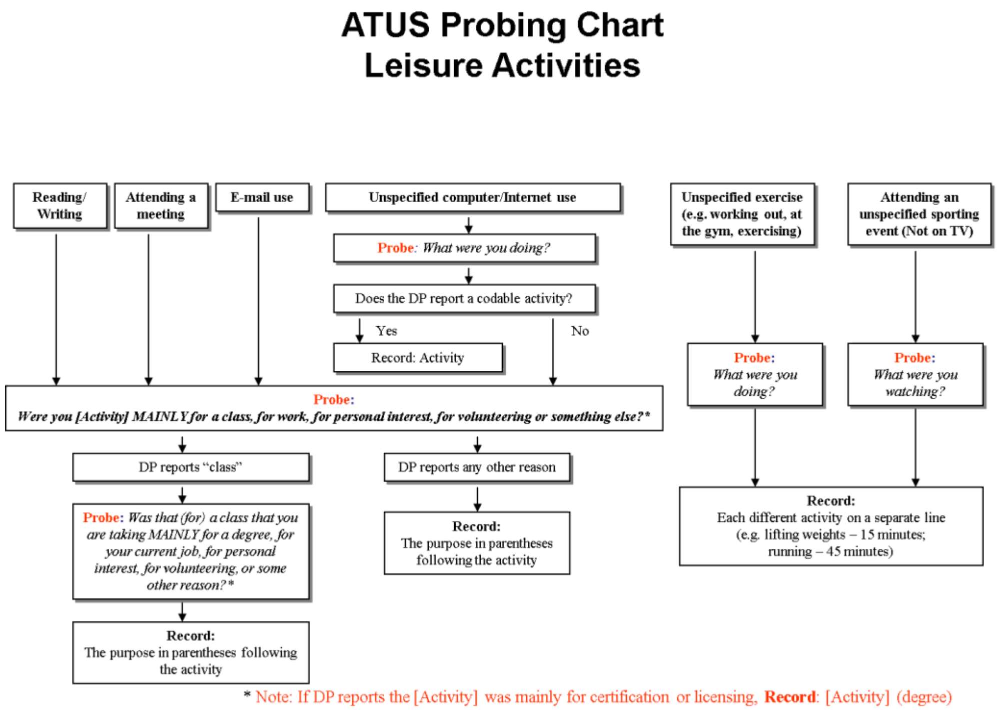

Figure 51 ATUS leisure activity probe chart

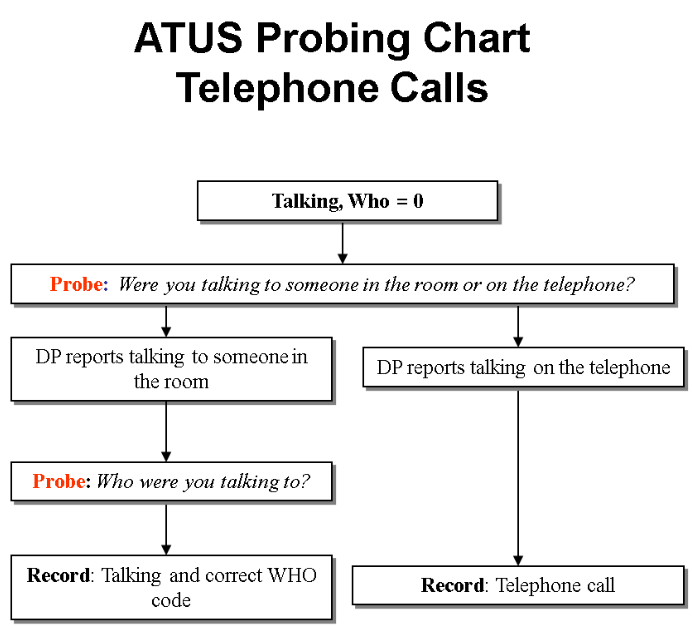

Figure 52 ATUS telephone calls activity probe chart

7.3 Post Interview Survey Questionnaire

1. The prompts were useful in this interview
    a. Strongly Disagree
    b. Disagree
    c. Neither Agree nor Disagree
    d. Agree
    e. Strongly Agree
2. Is there any reason this interview data should NOT be used?
    a. Yes
    b. No
3. ***If Q2 = Yes***; Why do you think the data should not be used?
    a. Respondent intentionally provided WRONG answers
    b. Respondent trying to provide RIGHT answers, but is unable to remember correctly
    c. Respondent deliberately reporting LONG duration activities
    d. Others ___
4. The instrument had a significant positive impact on the quality of the interview
    a. Strongly Disagree
    b. Disagree
    c. Neither Agree nor Disagree
    d. Agree
    e. Strongly Agree
5. The instrument had a significant negative impact on the quality of the interview
    a. Strongly Disagree
    b. Disagree
    c. Neither Agree nor Disagree
    d. Agree
    e. Strongly Agree
6. In my opinion, the impact that I (the interviewer) had on the quality of the interview as compared to the respondent was __
    a. Much smaller
    b. Smaller

c. Same

d. Larger

e. Much Larger

7. In my opinion, the impact that respondent had on the quality of the interview as compared to the instrument was __

   a. Much smaller
   b. Smaller
   c. Same
   d. Larger
   e. Much Larger

8. In my opinion, the impact that instrument had on the quality of the interview as compared to me (the interviewer) was __

   a. Much smaller
   b. Smaller
   c. Same
   d. Larger
   e. Much Larger

7.4 Web ATUS Phase 01 (With prompt panel)

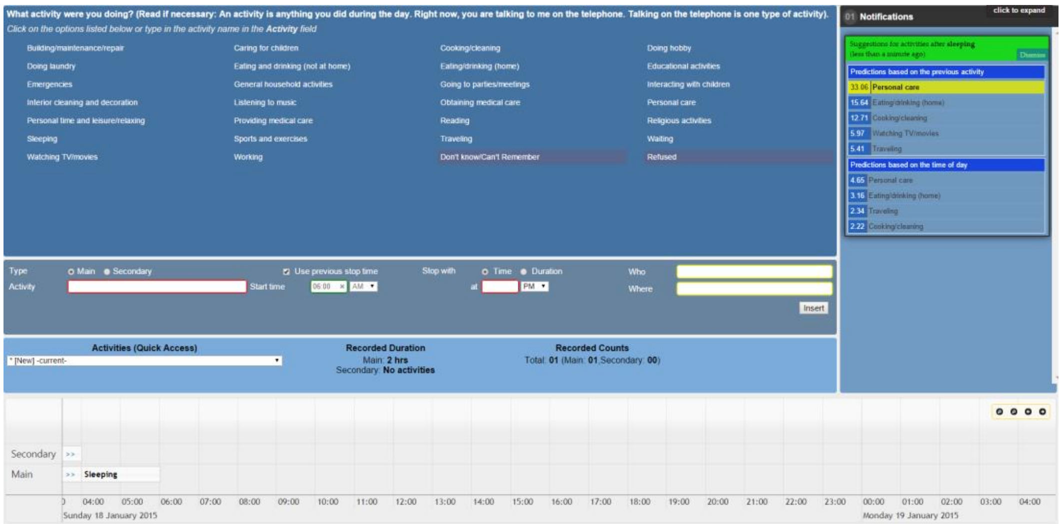

Figure 53 Phase 01 instrument with the prompt panel visible (top right panel)

7.5 Instrument Precode List

Waiting	Traveling
Personal care	Sleeping
Interacting with children	Cooking/cleaning
Doing hobby	Personal time and leisure/relaxing
Educational activities	Doing laundry
Religious activities	Obtaining medical care
Going to parties/meetings	Sports and exercises
Listening to music	Eating and drinking (not at home)
Emergencies	Caring for children
Building/maintenence/repair	Working
Reading	General household activities
Watching TV/movies	Providing medical care
Eating/drinking (home)	Refused
Interior cleaning and decoration	Don't know/Can't remember

Table 57 List of activities that were precodes in the Web ATUS instrument

7.6 Tier 3 Activity Transformation Table

Code	Tier 3 Activity	Mid Tier Activity	Mapped Concept
10101	Sleeping	Sleeping	Sleeping
10102	Sleeplessness	Sleeping	Sleeping
10201	Washing, dressing and grooming oneself	Personal care	Personal Care
10301	Health-related self-care	Personal care	Personal Care
80501	Using personal care services	Personal care	Professional Services
50101	Work - main job	Working	Working
50102	Work- other job(s)	Working	Working
50103	Security procedures related to work	Going through security	Working
50204	Security procedures as part of job	Going through security	Working
50405	Security procedures rel. to job search interviewing	Going through security	Working
70301	Security procedures rel. to consumer purchases	Going through security	Shopping
120405	Security procedures rel. to arts & entertainment (art entertainment)	Going through security	Outdoor Entertainment
140104	Security procedures rel. to religious & spiritual activities	Going through security	Religious
60104	Security procedures rel. to taking classes	Going through security	Education
80801	Security procedures rel. to professional personal svcs.	Going through security	Professional Services
100401	Security procedures rel. to govt svcs civic obligations	Going through security	Government
181801	Security procedures related to traveling	Going through security	Travelling

150801	Security procedures related to volunteer activities	Going through security	Volunteering
50104	Waiting associated with working	Waiting	Working
50205	Waiting associated with work-related activities	Waiting	Working
50305	Waiting associated with other income-generating activities	Waiting	Working
50404	Waiting associated with job search or interview	Waiting	Working
110201	Waiting associated w eating & drinking	Waiting	Food Eating & Preparation
90104	Waiting associated with using household services	Waiting	Household Activities
70105	Waiting associated with shopping	Waiting	Shopping
130301	Waiting related to playing sports or exercising	Waiting	Recreation - Outdoor
120501	Waiting assoc. w socializing & communicating	Waiting	Outdoor Entertainment
120502	Waiting assoc. w attending hosting social events	Waiting	Outdoor Entertainment
120503	Waiting associated with relaxing leisure	Waiting	Outdoor Entertainment
120504	Waiting associated with arts & entertainment	Waiting	Outdoor Entertainment
130302	Waiting related to attending sporting events	Waiting	Outdoor Entertainment
90202	Waiting associated with home main repair decor constr	Waiting	Maintenance & Repair Work
80403	Waiting associated with medical services	Waiting	Medical
30303	Waiting associated with hh children's health	Waiting	Medical

40303	Waiting associated with nonhh children's health	Waiting	Medical
140103	Waiting associated w religious & spiritual activities	Waiting	Religious
60103	Waiting associated with taking classes	Waiting	Education
60204	Waiting associated with extracurricular activities	Waiting	Education
60303	Waiting associated with research homework	Waiting	Education
60403	Waiting associated w admin. activities (education)	Waiting	Education
160201	Waiting associated with telephone calls	Waiting	Communication
30111	Waiting for with hh children	Waiting	Childcare
30204	Waiting associated with hh children's education	Waiting	Childcare
40111	Waiting for with nonhh children	Waiting	Childcare
40204	Waiting associated with nonhh children's education	Waiting	Childcare
80102	Waiting associated w purchasing childcare svcs	Waiting	Childcare
80702	Waiting associated with veterinary services	Waiting	Petcare
90302	Waiting associated with pet services	Waiting	Petcare
30405	Waiting associated with caring for household adults	Waiting	Adultcare
30504	Waiting associated with helping hh adults	Waiting	Adultcare
40405	Waiting associated with caring for nonhh adults	Waiting	Adultcare
40508	Waiting associated with helping nonhh adults	Waiting	Adultcare

80203	Waiting associated w banking financial services	Waiting	Professional Services
80302	Waiting associated with legal services	Waiting	Professional Services
80502	Waiting associated w personal care services	Waiting	Professional Services
80602	Waiting associated w purchasing selling real estate	Waiting	Professional Services
90402	Waiting associated with using lawn & garden services	Waiting	Professional Services
90502	Waiting associated with vehicle main. or repair svcs	Waiting	Professional Services
100304	Waiting associated with using government services	Waiting	Government
100305	Waiting associated with civic obligations & participation	Waiting	Government
150701	Waiting associated with volunteer activities	Waiting	Volunteering
50202	Eating and drinking as part of job	Going out to eat and drink	Working
10401	Personal Private activities	Personal time and leisure or relaxing	Personal Time
120301	Relaxing, thinking	Personal time and leisure or relaxing	Personal Time
50201	Socializing, relaxing and leisure as part of job	Going to parties or meetings	Working
120101	Socializing and communicating with others	Going to parties or meetings	Socializing
150501	Attending meetings, conferences & training	Going to parties or meetings	Socializing
120202	Attending meetings for personal interest (not volunteering)	Going to parties or meetings	Outdoor Entertainment

120201	Attending or hosting parties, receptions ceremonies	Going to parties or meetings	Socializing
30110	Attending hh children's events	Going to parties or meetings	Childcare
30202	Meetings and school conferences (hh children)	Going to parties or meetings	Childcare
40110	Attending nonhh children's events	Going to parties or meetings	Childcare
40202	Meetings and school conferences (nonhh children)	Going to parties or meetings	Childcare
50203	Sports and exercise as part of job	Sports and exercises	Working
50301	Income-generating hobbies, crafts and food	Income generating activities	Working
50302	Income-generating performances	Income generating activities	Working
50303	Income-generating services	Income generating activities	Working
50304	Income-generating rental property activities	Income generating activities	Working
80601	Activities rel. to purchasing selling real estate	Income generating activities	Professional Services
50401	Job search activities	Job searching and interviews	Working
50403	Job interviewing	Job searching and interviews	Working
20201	Food and drink preparation	Cooking and cleaning	Food Eating & Preparation
20202	Food presentation	Cooking and cleaning	Food Eating & Preparation
90102	Using meal preparation services	Cooking and cleaning	Food Eating & Preparation

150201	Food preparation, presentation, clean-up	Cooking and cleaning	Food Eating & Preparation
20203	Kitchen and food clean-up	Cooking and cleaning	Household Activities
40501	Housework, cooking & shopping assistance for nonhh adults	Cooking and cleaning	Household Activities
110101	Eating and drinking	Eating and Drinking at home	Food Eating & Preparation
20101	Interior cleaning	Interior cleaning and decoration	Household Activities
20104	Storing interior hh items, inc. food	Interior cleaning and decoration	Household Activities
20301	Interior arrangement, decoration & repairs	Interior cleaning and decoration	Household Activities
90101	Using interior cleaning services	Interior cleaning and decoration	Household Activities
90103	Using clothing repair and cleaning services	Interior cleaning and decoration	Household Activities
20103	Sewing, repairing & maintaining textiles	Interior cleaning and decoration	Maintenance & Repair Work
20303	Heating and cooling	Interior cleaning and decoration	Maintenance & Repair Work
20401	Exterior cleaning	Exterior cleaning and decoration	Maintenance & Repair Work
20402	Exterior repair, improvements & decoration	Exterior cleaning and decoration	Maintenance & Repair Work
20501	Lawn, garden and houseplant care	Lawn care and backyard activities	Maintenance & Repair Work
20502	Ponds, pools and hot tubs	Lawn care and backyard activities	Maintenance & Repair Work
90401	Using lawn and garden services	Lawn care and backyard activities	Professional Services

20701	Vehicle repair and maintenance (by self)	Building and maintenance and repair	Maintenance & Repair Work
20302	Building and repairing furniture	Building and maintenance and repair	Maintenance & Repair Work
20801	Appliance, tool and toy set-up repair & maintenance (by self)	Building and maintenance and repair	Maintenance & Repair Work
40502	House & lawn maintenance & repair assistance for nonhh adults	Building and maintenance and repair	Maintenance & Repair Work
40504	Vehicle & appliance maintenance repair assistance for nonhh adults	Building and maintenance and repair	Maintenance & Repair Work
90201	Using home maint repair decor construction svcs	Building and maintenance and repair	Maintenance & Repair Work
150301	Building houses, wildlife sites & other structures	Building and maintenance and repair	Maintenance & Repair Work
150302	Indoor & outdoor maintenance, repair & clean-up	Building and maintenance and repair	Maintenance & Repair Work
90501	Using vehicle maintenance or repair services	Building and maintenance and repair	Professional Services
20102	Laundry	Doing laundry	Household Activities
20902	Household & personal organization and planning	General household activities	Household Activities
20905	Home security	General household activities	Household Activities
40506	Household management & paperwork assistance for nonhh adults	General household activities	Household Activities
70101	Grocery shopping	Shopping	Shopping
70102	Purchasing gas	Shopping	Shopping
70103	Purchasing food (not groceries)	Shopping	Shopping
70104	Shopping, except groceries food and gas	Shopping	Shopping

70201	Comparison shopping	Shopping	Shopping
20901	Financial management	Finances management	Finances
40505	Financial management assistance for nonhh adults	Finances management	Finances
100103	Obtaining licenses & paying fines, fees taxes	Fees and taxes and licenses	Finances
120302	Tobacco and drug use	Tobacco and drug use	Recreation
130101	Doing aerobics	Aerobics and gymnastics	Recreation
130115	Doing gymnastics	Aerobics and gymnastics	Recreation
130103	Playing basketball	Playing basketball	Recreation
130105	Playing billiards	Playing billiards	Recreation
130107	Bowling	Playing bowling	Recreation
130109	Dancing	Dancing and other performances	Recreation
150401	Performing	Dancing and other performances	Recreation
130111	Fencing	Fencing	Recreation
130117	Playing hockey	Playing hockey	Recreation
130119	Participating in martial arts	Martial arts	Recreation
130120	Playing racquet sports	Playing racquet sports	Recreation
130122	Rollerblading	Rollerblading	Recreation
130124	Running	Running	Recreation
130126	Playing soccer	Playing soccer	Recreation
130127	Softball	Playing Softball	Recreation
130128	Using cardiovascular equipment	Gym and body training	Recreation
130133	Weightlifting strength training	Gym and body training	Recreation

130134	Working out, unspecified	Gym and body training	Recreation
130135	Wrestling	Gym and body training	Recreation
130130	Playing volleyball	Playing volleyball	Recreation
130131	Walking	Walking	Recreation
130132	Participating in water sports	Doing water sports and activities	Recreation
130106	Boating	Doing water sports and activities	Recreation - Outdoor
130136	Doing yoga	Doing yoga	Recreation
150103	Reading	Reading	Recreation
120312	Reading for personal interest	Reading	Indoor Entertainment
150105	Writing	Writing	Recreation
120313	Writing for personal interest	Writing	Indoor Entertainment
130102	Playing baseball	Playing baseball	Recreation - Outdoor
130104	Biking	Biking	Recreation - Outdoor
130108	Climbing, spelunking, caving	Hiking or climbing	Recreation - Outdoor
130116	Hiking	Hiking or climbing	Recreation - Outdoor
130110	Participating in equestrian sports	Equestrian and rodeo sports	Recreation - Outdoor
130121	Participating in rodeo competitions	Equestrian and rodeo sports	Recreation - Outdoor
130112	Fishing	Fishing or hunting	Recreation - Outdoor

130118	Hunting	Fishing or hunting	Recreation - Outdoor
130113	Playing football	Playing football	Recreation - Outdoor
130114	Golfing	Golfing	Recreation - Outdoor
130123	Playing rugby	Playing rugby	Recreation - Outdoor
130125	Skiing, ice skating snowboarding	Ice skating and skiing	Recreation - Outdoor
130129	Vehicle touring racing	Vehicle Racing	Recreation - Outdoor
120303	Television and movies (not religious)	Watching TV and movies	Indoor Entertainment
120304	Television (religious)	Watching TV and movies	Indoor Entertainment
120305	Listening to the radio	Listening to music	Indoor Entertainment
120306	Listening to playing music (not radio)	Listening to music	Indoor Entertainment
120307	Playing games	Playing video or computer games	Indoor Entertainment
120308	Computer use for leisure (exc. Games)	Recreational computer use	Indoor Entertainment
150101	Computer use	Recreational computer use	Personal Time
120309	Arts and crafts as a hobby	Doing hobby	Indoor Entertainment
120310	Collecting as a hobby	Doing hobby	Indoor Entertainment
120311	Hobbies except arts & crafts and collecting	Doing hobby	Indoor Entertainment

130232	Watching wrestling	Watching sports and games and activities	Indoor Entertainment
130201	Watching aerobics	Watching sports and games and activities	Outdoor Entertainment
130202	Watching baseball	Watching sports and games and activities	Outdoor Entertainment
130203	Watching basketball	Watching sports and games and activities	Outdoor Entertainment
130204	Watching biking	Watching sports and games and activities	Outdoor Entertainment
130205	Watching billiards	Watching sports and games and activities	Outdoor Entertainment
130206	Watching boating	Watching sports and games and activities	Outdoor Entertainment
130207	Watching bowling	Watching sports and games and activities	Outdoor Entertainment
130208	Watching climbing spelunking caving	Watching sports and games and activities	Outdoor Entertainment
130209	Watching dancing	Watching sports and games and activities	Outdoor Entertainment
130210	Watching equestrian sports	Watching sports and games and activities	Outdoor Entertainment
130211	Watching fencing	Watching sports and games and activities	Outdoor Entertainment
130212	Watching fishing	Watching sports and games and activities	Outdoor Entertainment
130213	Watching football	Watching sports and games and activities	Outdoor Entertainment
130214	Watching golfing	Watching sports and games and activities	Outdoor Entertainment
130215	Watching gymnastics	Watching sports and games and activities	Outdoor Entertainment

130216	Watching hockey	Watching sports and games and activities	Outdoor Entertainment
130217	Watching martial arts	Watching sports and games and activities	Outdoor Entertainment
130218	Watching racquet sports	Watching sports and games and activities	Outdoor Entertainment
130219	Watching rodeo competitions	Watching sports and games and activities	Outdoor Entertainment
130220	Watching rollerblading	Watching sports and games and activities	Outdoor Entertainment
130221	Watching rugby	Watching sports and games and activities	Outdoor Entertainment
130222	Watching running	Watching sports and games and activities	Outdoor Entertainment
130223	Watching skiing ice skating snowboarding	Watching sports and games and activities	Outdoor Entertainment
130224	Watching soccer	Watching sports and games and activities	Outdoor Entertainment
130225	Watching softball	Watching sports and games and activities	Outdoor Entertainment
130226	Watching vehicle touring racing	Watching sports and games and activities	Outdoor Entertainment
130227	Watching volleyball	Watching sports and games and activities	Outdoor Entertainment
130229	Watching water sports	Watching sports and games and activities	Outdoor Entertainment
120401	Attending performing arts	Attending galleries and museums and theaters	Outdoor Entertainment
120402	Attending museums	Attending galleries and museums and theaters	Outdoor Entertainment
120403	Attending movies film	Attending galleries and museums and theaters	Outdoor Entertainment

120404	Attending gambling establishments	Gambling	Outdoor Entertainment
30302	Obtaining medical care for hh children	Obtaining medical care	Medical
30404	Obtaining medical and care services for hh adult	Obtaining medical care	Medical
40302	Obtaining medical care for nonhh children	Obtaining medical care	Medical
40404	Obtaining medical and care services for nonhh adult	Obtaining medical care	Medical
80401	Using health and care services outside the home	Obtaining medical care	Medical
30301	Providing medical care to hh children	Providing medical care	Medical
30403	Providing medical care to hh adult	Providing medical care	Medical
40301	Providing medical care to nonhh children	Providing medical care	Medical
40403	Providing medical care to nonhh adult	Providing medical care	Medical
140101	Attending religious services	Religious activities	Religious
140102	Participation in religious practices	Religious activities	Religious
140105	Religious education activities: confirmation class) leading religious youth group	Religious activities	Religious
60101	Taking class for degree certification or licensure	Educational activities	Education
60102	Taking class for personal interest	Educational activities	Education
60301	Research homework for class for degree certification or licensure	Educational activities	Education
60302	Research homework for class for pers. interest	Educational activities	Education
60401	Administrative activities: class for degree certification or licensure	Educational activities	Education
60402	Administrative activities: class for personal interest	Educational activities	Education

150204	Teaching, leading counseling mentoring	Educational activities	Education
30203	Home schooling of hh children	Educational activities	Childcare
30201	Homework (hh children)	Educational activities	Childcare
40201	Homework (nonhh children)	Educational activities	Childcare
40203	Home schooling of nonhh children	Educational activities	Childcare
60201	Extracurricular club activities	Extracurricular activities	Education
60202	Extracurricular music & performance activities	Extracurricular activities	Education
60203	Extracurricular student government activities	Extracurricular activities	Education
20903	HH & personal mail & messages (except e-mail)	Mailing and messaging activities	Communication
20904	HH & personal e-mail and messages	Mailing and messaging activities	Communication
160101	Telephone calls to from family members	Talking on the telephone	Communication
160102	Telephone calls to from friends, neighbors, or acquaintances	Talking on the telephone	Communication
160103	Telephone calls to from education services providers	Talking on the telephone	Communication
160104	Telephone calls to from salespeople	Talking on the telephone	Communication
160105	Telephone calls to from professional or personal care svcs providers	Talking on the telephone	Communication
160106	Telephone calls to from household services providers	Talking on the telephone	Communication
160107	Telephone calls to from paid child or adult care providers	Talking on the telephone	Communication
160108	Telephone calls to from government officials	Talking on the telephone	Communication
150104	Telephone calls (except hotline counseling)	Talking on the telephone	Communication

10501	Personal emergencies	Emergencies	Personal Time
30101	Physical care for hh children	Caring for children	Childcare
30109	Looking after hh children (as a primary activity)	Caring for children	Childcare
40101	Physical care for nonhh children	Caring for children	Childcare
40109	Looking after nonhh children (as primary activity)	Caring for children	Childcare
80101	Using paid childcare services	Caring for children	Childcare
30102	Reading to with hh children	Interacting with children	Childcare
30103	Playing with hh children, not sports	Interacting with children	Childcare
30104	Arts and crafts with hh children	Interacting with children	Childcare
30105	Playing sports with hh children	Interacting with children	Childcare
30106	Talking with listening to hh children	Interacting with children	Childcare
30108	Organization & planning for hh children	Interacting with children	Childcare
40102	Reading to with nonhh children	Interacting with children	Childcare
40103	Playing with nonhh children, not sports	Interacting with children	Childcare
40104	Arts and crafts with nonhh children	Interacting with children	Childcare
40105	Playing sports with nonhh children	Interacting with children	Childcare
40106	Talking with listening to nonhh children	Interacting with children	Childcare
40108	Organization & planning for nonhh children	Interacting with children	Childcare
20601	Care for animals and pets (not veterinary care)	Petcare and related	Petcare
20602	Walking exercising playing with animals	Petcare and related	Petcare
40503	Animal & pet care assistance for nonhh adults	Petcare and related	Petcare
80701	Using veterinary services	Petcare and related	Petcare

90301	Using pet services	Petcare and related	Petcare
30401	Physical care for hh adults	Caring for other adults	Adultcare
30402	Looking after hh adult (as a primary activity)	Caring for other adults	Adultcare
30501	Helping hh adults	Caring for other adults	Adultcare
30502	Organization & planning for hh adults	Caring for other adults	Adultcare
40401	Physical care for nonhh adults	Caring for other adults	Adultcare
40402	Looking after nonhh adult (as a primary activity)	Caring for other adults	Adultcare
80402	Using in-home health and care services	Caring for other adults	Adultcare
150203	Providing care	Caring for other adults	Adultcare
80201	Banking	Banking and financial activities	Professional Services
80202	Using other financial services	Banking and financial activities	Professional Services
80301	Using legal services	Legal activities	Professional Services
100101	Using police and fire services	Public and Emergency services	Professional Services
100102	Using social services	Public and Emergency services	Professional Services
150601	Public health activities	Public and Emergency services	Volunteering
150602	Public safety activities	Public and Emergency services	Volunteering
100201	Civic obligations & participation	Performing civic duties	Government
150102	Organizing and preparing	Volunteer activities	Volunteering
150402	Serving at volunteer events & cultural activities	Volunteer activities	Volunteering

150106	Fundraising	Charity and fundraising	Volunteering
150202	Collecting & delivering clothing & other goods	Charity and fundraising	Volunteering
30112	Picking up dropping off hh children	Traveling	Childcare
40112	Dropping off picking up nonhh children	Traveling	Childcare
30503	Picking up dropping off hh adult	Traveling	Adultcare
40507	Picking up dropping off nonhh adult	Traveling	Adultcare
180101	Travel related to personal care	Traveling	Travelling
180201	Travel related to housework	Traveling	Travelling
180202	Travel related to food & drink prep. clean-up & presentation	Traveling	Travelling
180203	Travel related to interior maintenance repair & decoration	Traveling	Travelling
180204	Travel related to exterior maintenance repair & decoration	Traveling	Travelling
180205	Travel related to lawn garden and houseplant care	Traveling	Travelling
180206	Travel related to care for animals and pets (not vet care)	Traveling	Travelling
180207	Travel related to vehicle care & maintenance (by self)	Traveling	Travelling
180208	Travel related to appliance tool and toy set-up repair & maintenance (by self)	Traveling	Travelling
180209	Travel related to household management	Traveling	Travelling
180301	Travel related to caring for & helping hh children	Traveling	Travelling
180302	Travel related to hh children's education	Traveling	Travelling
180303	Travel related to hh children's health	Traveling	Travelling

180304	Travel related to caring for hh adults	Traveling	Travelling
180305	Travel related to helping hh adults	Traveling	Travelling
180401	Travel related to caring for and helping nonhh children	Traveling	Travelling
180402	Travel related to nonhh children's education	Traveling	Travelling
180403	Travel related to nonhh children's health	Traveling	Travelling
180404	Travel related to caring for nonhh adults	Traveling	Travelling
180405	Travel related to helping nonhh adults	Traveling	Travelling
180501	Travel related to working	Traveling	Travelling
180502	Travel related to work-related activities	Traveling	Travelling
180503	Travel related to income-generating activities	Traveling	Travelling
180504	Travel related to job search & interviewing	Traveling	Travelling
180601	Travel related to taking class	Traveling	Travelling
180602	Travel related to extracurricular activities (ex. Sports)	Traveling	Travelling
180603	Travel related to research homework	Traveling	Travelling
180604	Travel related to registration administrative activities	Traveling	Travelling
180701	Travel related to grocery shopping	Traveling	Travelling
180702	Travel related to purchasing gas	Traveling	Travelling
180703	Travel related to purchasing food (not groceries)	Traveling	Travelling
180704	Travel related to shopping ex groceries food and gas	Traveling	Travelling
180801	Travel related to using childcare services	Traveling	Travelling

180802	Travel related to using financial services and banking	Traveling	Travelling
180803	Travel related to using legal services	Traveling	Travelling
180804	Travel related to using medical services	Traveling	Travelling
180805	Travel related to using personal care services	Traveling	Travelling
180806	Travel related to using real estate services	Traveling	Travelling
180807	Travel related to using veterinary services	Traveling	Travelling
180901	Travel related to using household services	Traveling	Travelling
180902	Travel related to using home main. repair decor. construction svcs	Traveling	Travelling
180903	Travel related to using pet services (not vet)	Traveling	Travelling
180904	Travel related to using lawn and garden services	Traveling	Travelling
180905	Travel related to using vehicle maintenance & repair services	Traveling	Travelling
181001	Travel related to using government services	Traveling	Travelling
181002	Travel related to civic obligations & participation	Traveling	Travelling
181101	Travel related to eating and drinking	Traveling	Travelling
181201	Travel related to socializing and communicating	Traveling	Travelling
181202	Travel related to attending or hosting social events	Traveling	Travelling
181203	Travel related to relaxing and leisure	Traveling	Travelling
181204	Travel related to arts and entertainment	Traveling	Travelling
181205	Travel as a form of entertainment	Traveling	Travelling
181301	Travel related to participating in sports exercise recreation	Traveling	Travelling

181302	Travel related to attending sporting recreational events	Traveling	Travelling
181401	Travel related to religious spiritual practices	Traveling	Travelling
181501	Travel related to volunteering	Traveling	Travelling
181601	Travel related to phone calls	Traveling	Travelling

Table 58 ATUS Tier 3 activities to MID tier activity and mapped concepts translation

Chapter 8: References

Aamodt, Agnar, and Enric Plaza. "Case-based reasoning: Foundational issues, methodological variations, and system approaches." *AI communications* 7.1 (1994): 39-59.

Aamodt, A., (1995). Knowledge Acquisition and Learning by Experience – The role of Case-Specific Knowledge. *Machine learning and knowledge acquisition: integrated approaches*. London, UK: Academic Press Ltd.

Adams R J. (1990), A System to Improve the Availability of Information for Library Managers. In Burton P F (editor). *MS-DOS Software For Library And Information Applications*, Gower Publishing 1990 ISBN 0 566 03617 7 pp123-135, Emerald Back-files 2007

Adomavicius, Gediminas, and Alexander Tuzhilin. "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions." *Knowledge and Data Engineering, IEEE Transactions* on 17.6 (2005): 734-749.

Adomavicius, Gediminas, et al. "Incorporating contextual information in recommender systems using a multidimensional approach." *ACM Transactions on Information Systems* (TOIS) 23.1 (2005): 103-145.

Apolloni, Bruno, Giacomo Zamponi, and Anna Maria Zanaboni. "Learning fuzzy decision trees." *Neural Networks* 11.5 (1998): 885-895.

Andreica M, "Overview of Modelling the Risk Assessment Process in Applied Project Management," *Metalurgia International*, vol. 14, pp. 139-144, 2009.

Andrews, Dorine, Blair Nonnecke, and Jennifer Preece. "Electronic survey methodology: A case study in reaching hard-to-involve Internet users.*" International journal of human-computer interaction* 16.2 (2003): 185-210.

Armstrong, J.S., *Principles of Forecasting—A Handbook for Researchers and Practitioners*. Kluwer Academic, 2001.

Backor, Kristen, Saar Golde, and Norman Nie. "Estimating survey fatigue in time use study." *International Association for Time Use Research Conference,* Washington, DC. 2007.

Balabanović, Marko, and Yoav Shoham. "Fab: content-based, collaborative recommendation." *Communications of the ACM* 40.3 (1997): 66-72.

Bechwati, Nada Nasr, and Lan Xia. "Do computers sweat? The impact of perceived effort of online decision aids on consumers' satisfaction with the decision process." *Journal of Consumer Psychology* 13.1 (2003): 139-148.

Belkin, Nicholas J., and W. Bruce Croft. "Information filtering and information retrieval: two sides of the same coin?." Communications of the ACM 35.12 (1992): 29-38.

Bernoulli, Daniel. "Exposition of a new theory on the measurement of risk."*Econometrica: Journal of the Econometric Society* (1954): 23-36.

Bilgic, Mustafa, and Raymond J. Mooney. "Explaining recommendations: Satisfaction vs. promotion." *Beyond Personalization Workshop, IUI*. Vol. 5. 2005.

Bolger, Niall, Angelina Davis, and Eshkol Rafaeli. "Diary methods: Capturing life as it is lived." *Annual review of psychology* 54.1 (2003): 579-616.

Cetintas, Suleyman, et al. "Automatic detection of off-task behaviors in intelligent tutoring systems with machine learning techniques." *Learning Technologies, IEEE Transactions* on 3.3 (2010): 228-236.

Cobanoglu, Cihan, Bill Warde, and Patrick J. Moreo. "A comparison of mail, fax and web-based survey methods." International journal of market research 43.4 (2001): 441-452.

Conrad, Frederick G., et al. "Comprehension and engagement in survey interviews with virtual agents." *Frontiers in psychology* 6 (2015).

Couper, Mick P. "Review: Web surveys: A review of issues and approaches." *Public opinion quarterly* (2000): 464-494.

Crosbie, Tracey. "Using activity diaries: Some methodological lessons." *Journal of Research Practice* 2.1 (2006): 1.

Currivan, Douglas B., et al. "Does telephone audio computer-assisted self-interviewing improve the accuracy of prevalence estimates of youth smoking? Evidence from the UMass Tobacco Study." Public Opinion Quarterly 68.4 (2004): 542-564.

Damodaran, A. "Risk management instruments for debt driven conservation efforts: The case of India's Project Tiger." *Ecological Economics* 68.3 (2009): 625-633.

Dawood, Imad, and Mustafa Alshawi. "Decision Support Systems (DSS) Model for the Housing Industry." *Developments in eSystems Engineering (DESE), 2009 Second International Conference on*. IEEE, 2009.

Dey, Prasanta Kumar. "Project risk management: a combined analytic hierarchy process and decision tree approach." *Cost Engineering* 44.3 (2002): 13-27.

Dillman, Don A., and Dennis K. Bowker. "The web questionnaire challenge to survey methodologists." *Bemad Batinic, Ulf-Dietrich Reips, Michael Bosnjak & Andreas Werner: Online Social Sciences. Seattle: Hogrefe & Huber* (2001): 53-71.

Efstathiou, Nikolaos, Jamal Ameen, and Anne-Marie Coll. "Healthcare providers' priorities for cancer care: a Delphi study in Greece." *European Journal of Oncology Nursing* 11.2 (2007): 141-150.

Fishburn, Peter C. *Utility theory for decision making*. No. RAC-R-105. RESEARCH ANALYSIS CORP MCLEAN VA, 1970.

Ford, K., Bradshaw, J., Adams-Webber, J., & Agnew, N., (1993). Knowledge Acquisition as a constructive modeling activity. *International Journal of Intelligent Systems*, 8, 9-31.

Forsythe, B. "Assessing data quality: How well did our approach work."Conference on Time Use, Non-Market Work and Family Well-Being. Washington, DC: *Bureau of Labor Statistics and MacArthur Network on the Family and the Economy*. 1997.

Gelly, Sylvain, and David Silver. "Combining online and offline knowledge in UCT." *Proceedings of the 24th international conference on Machine learning*. ACM, 2007.

Graesser, Arthur C., et al. "Survey interviews with new communication technologies: Synthesis and future opportunities." *Envisioning the survey interview of the future* (2008): 267-284.

Gretzel, Ulrike, and Daniel R. Fesenmaier. "*Persuasion in recommender systems.*" International Journal of Electronic Commerce 11.2 (2006): 81-100.

Hamilton, Diane Matteo, and Steven Breslawski. "Knowledge acquisition for multiple site, related domain expert systems: Delphi process and application."*Expert Systems with Applications* 11.3 (1996): 377-389.

Häubl, Gerald, and Kyle B. Murray. "Preference construction and persistence in digital marketplaces: The role of electronic recommendation agents.*" Journal of Consumer Psychology* 13.1-2 (2003): 75-91.

Henning, G. (1987) A guide to language testing. Cambridge, Mass.: *Newbury House*

Hill, William C., et al. "Edit wear and read wear." *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 1992.

Horrigan, Michael, and Diane Herz. "A study in the process of planning, designing and executing a survey program: the BLS American time-use survey." *Contributions to Economic Analysis* 271 (2004): 317-350.

Joinson, Adam N. "Self‐disclosure in computer‐mediated communication: The role of self‐awareness and visual anonymity." *European Journal of Social Psychology* 31.2 (2001): 177-192.

Kamba, Tomonari, Hidekazu Sakagami, and Yoshiyuki Koseki. "ANATAGONOMY: a personalized newspaper on the World Wide Web."*International Journal of Human-Computer Studies* 46.6 (1997): 789-803.

Khashei, Mehdi, Seyed Reza Hejazi, and Mehdi Bijari. "A new hybrid artificial neural networks and fuzzy regression model for time series forecasting.*" Fuzzy Sets and Systems* 159.7 (2008): 769-786.

Kite, Jared J. A Flexible Framework for Knowledge Engineering and Automation of an Adaptive Conversational Case-retrieval System. Diss. *University of Nebraska--Lincoln*, 2007.

Kolodner, Janet. *Case-based reasoning*. Morgan Kaufmann, 2014.

Kramer, Thomas. *The effect of preference measurement on preference construction and responses to customized offers*. Diss. Stanford University, 2003.

Krosnick, Jon A. "Response strategies for coping with the cognitive demands of attitude measures in surveys." *Applied cognitive psychology* 5.3 (1991): 213-236.

Kruger, Justin, et al. "The effort heuristic." *Journal of Experimental Social Psychology* 40.1 (2004): 91-98.

Leake, David B. "CBR in context: the present and future." *Case-Based Reasoning, Experiences, Lessons & Future Directions* (1996): 1-30.

Lilien, G.L., Kotler, P, and Moorthy, K.S., Marketing Models. *Prentice Hall*, 1992.

Linacre, John Michael. "Computer-adaptive testing: A methodology whose time has come." Chae, S.-Kang, U.–Jeon, E.–Linacre, JM (eds.): *Development of Computerised Middle School Achievement Tests, MESA Research Memorandum* 69 (2000).

Litwin, Mark S. *How to measure survey reliability and validity*. Vol. 7. Sage Publications, 1995.

Malhotra, Yogesh, and D. F. Galleta. "Role of commitment and motivation in knowledge management systems implementation: Theory, conceptualization, and measurement of antecedents of success." System Sciences, 2003. *Proceedings of the 36th Annual Hawaii International Conference on. IEEE*, 2003.

Mandel, Naomi, and Eric J. Johnson. "When web pages influence choice: Effects of visual primes on experts and novices." *Journal of Consumer Research* 29.2 (2002): 235-245.

de Meaux, Juliette, and Maarten Koornneef. "The cause and consequences of natural variation: the genome era takes off!." *Current opinion in plant biology*11.2 (2008): 99-102.

Minnen, Joeri, et al. "Modular Online Time Use Survey (MOTUS)–Translating an existing method in the 21 st century." *electronic International Journal of Time Use Research* (2014): 73.

Morita, Masahiro, and Yoichi Shinoda. "Information filtering based on user behavior analysis and best match text retrieval." *Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval*. Springer-Verlag New York, Inc., 1994.

Morwitz, Vicki G., Eric Johnson, and David Schmittlein. "Does measuring intent change behavior?." *Journal of consumer research* (1993): 46-61.

Murray, Kyle B., and Gerald Häubl. "Processes of preference construction in agent-assisted online shopping." *Online Consumer Psychology: Understanding and Influencing Consumer Behavior in the Virtual World* (2005): 246.

Murthi, B.P.S. & Sarkar, S, "The Role of the Management Sciences in Research on Personalization," *Management Science, vol. 49, no. 10*, pp. 1344-1362, 2003.

Nass, Clifford, and Youngme Moon. "Machines and mindlessness: Social responses to computers." *Journal of social issues* 56.1 (2000): 81-103.

O'connor, Mark, et al. "PolyLens: a recommender system for groups of users."*ECSCW 2001*. Springer Netherlands, 2001.

Packer, Katherine H., and Dagobert Soergel. "The importance of SDI for current awareness in fields with severe scatter of information." *Journal of the American Society for Information Science* 30.3 (1979): 125-135.

Pannell, D.J. (1997). Sensitivity Analysis of Normative Economic Models: Theoretical Framework and Practical Strategies. Agricultural Economics 16: 139-152, at URL: http://cyllene.uwa.edu.au/~dpannell/dpap971f.htm

Power, D. J., (2000) "A Brief History of Spreadsheets". Version 3.3, March 11, 2000, URL: http://dssresources.com/history/sshistory.html Accessed (05/28/2014).

Powell, M.J.D, *Approximation Theory and Methods*. Cambridge Univ. Press, 1981.

Kaelbling, Leslie Pack, Michael L. Littman, and Andrew W. Moore. "Reinforcement learning: A survey." *arXiv preprint cs/9605103* (1996).

Ramos, Carlos, Claude Frasson, and Sowmya Ramachandran. "Introduction to the special issue on real world applications of intelligent tutoring systems."*Learning Technologies, IEEE Transactions* on 2.2 (2009): 62-63.

Rich, E, "User Modeling via Stereotypes," *Cognitive Science, vol. 3, no. 4,* pp. 329-354, 1979.

Reeves, Byron, and Clifford Nass. *How people treat computers, television, and new media like real people and places*. CSLI Publications and Cambridge University press, 1996.

Salton, G, *Automatic Text Processing*. Addison-Wesley, 1989.

Schank, Roger C. "Explanation patterns: Understanding mechanically and creatively." (1986).

Seo, Young-Woo, and Byoung-Tak Zhang. "Learning user's preferences by analyzing Web-browsing behaviors." *Proceedings of the fourth international conference on Autonomous agents*. ACM, 2000.

Shanks, J. Merrill. "The Current Status of Computer-Assisted Telephone Interviewing Recent Progess and Future Prospects." *Sociological Methods & Research* 12.2 (1983): 119-142.

Simonson, Itamar. "Determinants of customers' responses to customized offers: Conceptual framework and research propositions." *Journal of Marketing* 69.1 (2005): 32-45.

Stinson, Linda L. "Measuring how people spend their time: a time-use survey design." *Monthly Lab*. Rev. 122 (1999): 12.

Stopher, Peter, Camden FitzGerald, and Min Xu. "Assessing the accuracy of the Sydney Household Travel Survey with GPS." *Transportation* 34.6 (2007): 723-741.

Studer R., Benjamins, V.R., & Fensel, D., (1998). Knowledge Engineering: Principles and Methods, *Data Knowledge Engineering*, 25, 161-197.

Ungar, Lyle H., and Dean P. Foster. "Clustering methods for collaborative filtering." *AAAI Workshop on Recommendation Systems*. Vol. 1. 1998.

Von Neumann, John, and Oskar Morgenstern. "Theory of games and economic behavior (2d rev." (1947).

Wright, Kevin B. "Researching Internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services." *Journal of Computer‐Mediated Communication* 10.3 (2005).

Xiaocong, He, and Kang Ling. "A risk management decision support system for project management based on bayesian network." *Information Management and Engineering (ICIME), 2010 The 2nd IEEE International Conference on*. IEEE, 2010.

Yates, R. Baeza and Neto, B. Ribeiro, *Modern Information Retrieval*. Addison-Wesley, 1999.

Yegnanarayana, B. *Artificial neural networks*. PHI Learning Pvt. Ltd., 2009.