

TECHNICAL AND USABILITY TESTING OF A MEDICAL INTERACTION WORKGROUP APPLICATION SUPPORT SYSTEM (MIWASS)

James A. Rodger, Indiana University of Pennsylvania, jrodger@iup.edu

ABSTRACT

The Medical Interaction Workgroup Application Support System (MIWASS) was run on portable handheld Personal Digital Assistant (PDA) devices. The MIWASS dental and veterinary software modules were evaluated by users from a group of subjects attending national and international conferences in the United States. The end users agreed to use the application in exchange for a chance to win a PDA mobile device. The user group consisted of 36 dentists and 36 veterinarians.

Keywords: Usability, Decision Support, Task Technology Fit, Use

INTRODUCTION

The study began in 2006 to explore the implementation of the portable platform for electronic collection of operational-level dental and veterinary information at the point-of-care. During the first year of this study, we developed the MIWASS dental and veterinary modules. These modules, based on requirements derived from Army dental and veterinary forms, show promise in confirming hypotheses that the platform can improve dental and veterinary care. These improvements would result from an increase in the accuracy of information necessary for medical situational awareness.

The MIWASS dental and veterinary modules will minimize errors in point-of-care data entry resulting from hand transcription. The use of these modules, integrated with a platform installed on a handheld device, will also increase flexibility in data collection during fieldwork. These hypotheses arise from the findings documented in previous studies and articles. According to one such article, as advances in information technology increase the potential to collect, store, and share information, entire fields of research dedicated to the integration of technology with practice have developed. In the medical context, medical informatics has developed as a legitimate field of study, aimed at finding creative solutions to digitizing and simplifying medical

information collection, storage, and presentation. The medical informatics field “deals with the acquisition, compilation, storage, and appropriate distribution of clinical or other health data” (Schomer 1999). Much of the drive to introduce information technology into the medical field stems from numerous studies that have found varying degrees of human error in medical settings, and the realization of potential inefficiencies inherent in current methods.

Though estimates of impact have varied from study to study, what is clear is that medical errors have a significant and detrimental effect on patient care. Medical mistakes were the focus of a widely-noted 1999 national report that estimated 44,000 to 98,000 Americans die each year because of errors and low-quality care (O’Leary 2006). Another study sampled pediatric immunization records in a major hospital system during a 12-month period (Wilton and Pennisi 1994). This facility used hand-written documentation, which was later transcribed on-site by a medical transcriptionist. The authors found that 22.5% of the records examined were not up-to-date. Additional examination of these cases revealed that at least 10.2% (214 records) had at least one transcription error. Moreover, 72 charts had immunization information documented in sections that were not routinely transcribed (Wilton and Pennisi 1994).

A 2003 British study provided journals to ten general medical practices for a two-week period. During this time, healthcare workers at these facilities were instructed to document any observed errors, defined as “an event that was not completed as intended and/or meant that work was disrupted in some way” (Rubin, Chinn, and Richardson 2003). Of relevance to this research is the preponderance of prescription and communication errors these respondents discovered. Approximately 42% (397) of the recorded errors related to prescriptions. Furthermore, 30% of the errors were communication errors, with missing case notes most prominent.

Medical transcription errors have been identified as the eighth leading cause of death in the United States

and can lead to iatrogenic morbidity and mortality (Schomer 1999). These errors can occur at the point of service as handwriting errors, or in the data conversion phase as transcription errors. Examples of poor handwriting in a medical setting have led to the prescription of inappropriate medications and synergistic drug interactions (Rodger, Pendharkar, and Paper 1999). Exacerbating this problem is the dereliction of reporting duty by medical staff. A survey of thirteen US hospitals found that over 50% (850) of medical staff witnessed a medical mistake made by a colleague. However, less than 10% discussed the error with the colleague (Jablow 2005).

Realization of the prevalence of medical errors, or adverse events, has inspired lawmakers and researchers to propose reform. In particular, lawmakers have pushed hospitals to introduce computerized physician-order entry (CPOE) systems to aid in medication prescription entry (Davies 2005). Despite the increased interest in medical informatics, a 2005 study asserted that only five to nine percent of US hospitals have implemented some form of CPOE (Koppel, et al. 2005).

While some of this delay in transition can be attributed to variations in organizational rates of technology acquisition and implementation, the lack of widespread acceptance of CPOE and other medical point-of-care technologies underlies a greater concern. While there is a body of literature comparing the various methods of data collection and entry, the relative benefits of each approach are often unclear in some studies. Conflicting studies have left open the possibility that direct physician input of medical information may introduce greater error, relative to handwritten and transcribed records.

For example, a 2005 study found that a CPOE system introduced at the Hospital of University of Pennsylvania increased the probability of twenty-two types of “situations,” or errors. These errors principally involved failure due to data and system fragmentation, and human-machine interface errors (Koppel, et al. 2005). However, this study is limited insofar as it compared the hospital’s current CPOE system against a previous system, and not handwritten-transcription methods (Davies 2005).

Admittedly, a meta-evaluation by Hogan and Wagner revealed that little research had compared direct physician data entry systems against the paper standard (Hogan and Wagner 1997). These authors did, however, acknowledge that structured data forms significantly improved accuracy relative to unstructured forms. Kuhn, et al. (1992), supports

this. Even fewer studies have compared the error rates of handheld point-of-care devices versus the central terminal standard. Despite the inconclusive debate, the landmark Institute of Medicine (IoM) report compels the medical field to adopt point-of-care direct physician input (Dick and Steen 1991).

Creating a point-of-care system requires the entry of information during the physician-patient interaction. Therefore, terminal-based computing does not seem to fit the IoM requirements. However, the use of PDAs offers great promise in terms of mobility, convenience, and accuracy.

The MIWASS team has developed PDA modules capable of capturing, storing, and processing data. Improved data acquisition and transmission capabilities will allow connectivity to other systems. The MIWASS dental and veterinary modules have been programmed with preformatted, structured forms. We submit that this innovation will improve data quality, efficiency, and usability.

The PDA-based modules developed for this effort will add the necessary flexibility required for fieldwork, which will be a boon to military veterinarians. In fact, the nature of fieldwork necessitates the flexibility provided by PDAs, which paper-based and laptop-based technologies lack. Pascoe, Ryan, and Morris successfully introduced a modified PDA for use in an ecology study in Kenya (2000). From this software trial, these authors summarized four characteristics of fieldworkers’ work patterns that necessitate PDAs. These are dynamic use, limited attention, high-speed interaction, and context dependence. For the dental component, one can assume that dynamic use and limited attention apply.

Prior to software development, several limitations and benefits of handheld devices were considered. There is a wealth of case studies in which healthcare practitioners were provided with handheld devices. Their criticisms and positive comments are similar in each study. In Chiang, et al., ten medical residents tested a handheld device that was “hot-synched” to a patient records database (2003). Using structured forms, residents were able to enter patient information directly into the handheld device, which was then downloaded to a central workstation via a cable link. The chief criticisms among this cohort were slow CPU speed, particularly with low-end PDAs, and a “cumbersome user interface” (Chiang, et al. 2003). Another study echoed these criticisms. Respondents indicated that the screen size was too small for text-intensive purposes, and the system was

unstable at times (Carroll, Saluja, and Tarczy-Hornoch 2002).

Finally, a 2001 study of 72 nurses in a Taipei hospital revealed that a statistically significant number of respondents found the PDA application easier to operate than the traditional terminal workstation (Chang, et al. 2003). However, they also cited interface concerns. This study went a step further by introducing technology acceptance questions, based on the Technology Acceptance Model (TAM) (Davis 1989). Here, the authors found that nurses were statistically less willing to accept the PDA technology than the terminal technology (Chang, et al. 2003). However, this study was limited insofar as it only provided nurses four four-hour sessions with the PDA. As Hubona and Burton-Jones indicate, there is a moderate direct effect of time since first use on beliefs about usefulness (2002).

A review of the literature on current PDA use in the medical field reveals untapped potential. To date, a majority of previous handheld device use by physicians has been limited to decision support. Organizations such as Handheldmed, Inc. (www.handheldmed.com), Lippincott Williams & Wilkins (www.lww.com/stores/pda/pdalist.html), ANMLDR (<http://www.anmldr.com/>), and the University of California–Davis School of Veterinary Medicine (<http://vetpda.ucdavis.edu/Projects/VetPDA.cfm>) have developed applications to be run on a number of handheld platforms. However, these applications are generally limited to reference materials, dosage guides, and differential diagnostic assistance.

We believe that by integrating standardized forms into the platform, a robust, quality-driven system will replace redundant and inefficient paper data collection methods.

METHODOLOGY

We set out to demonstrate the usability and ease of use of the MIWASS dental and veterinary software modules by conducting usability testing as part of the study. This usability testing included the distribution of a survey to trained users of the MIWASS dental and veterinary modules. This survey incorporated a six-point Likert scale for measurement purposes. The survey featured 16 closed-ended questions.

Our measurement method for user evaluations is supported by the literature (Folmer and Bosch 2004; Chen, et al. 2004; Koppel, et al. 2005). In an article discussing methods for measuring usability, Folmer

wrote that closed-ended surveys/questionnaires are an excellent way to measure user usability. He explained that questionnaires are widely used and reliable. Folmer lists the following as reliable questionnaires: Questionnaire for User Interface Satisfaction (QUIS), Perceived Usefulness and Ease of Use (PUEU), and User Satisfaction Instrument (USI) (Folmer and Bosch 2004). The MIWASS team utilized these questionnaires as references for survey development.

The basic rationale of the MIWASS team is that a survey/questionnaire administered to users will provide important feedback about the dental and veterinary modules developed for the PDA platform. After gathering such feedback, MIWASS software developers will consider using the survey results as the basis for improvements to the MIWASS dental and veterinary modules. This could enhance the quality, efficiency, and usability of the dental and veterinary modules.

In addition to using a survey to measure user satisfaction with the MIWASS dental and veterinary modules, the MIWASS team embedded code in the software to gain insights about the efficiency of the modules. This embedded code measured aspects such as the length of time users took to complete tasks. Hilbert provided justification for the use of embedded coding. He noted that “built-in” codes are an excellent way to provide summary statistics pertaining to user behavior (Hilbert and Redmiles 2000). We applied the results of the embedded code and survey to measure the quality, efficiency, and usability of the existing MIWASS dental and veterinary modules.

QUALITY

There is general agreement in the literature concerning the definition of quality. Quality is considered to be a multidimensional concept, encompassing accuracy, completeness, consistency, and timeliness (Wand and Wang 1996). These dimensions are defined as follows:

- Accuracy is most commonly defined as correctness (Wand and Wang 1996). In other words, accuracy refers to the precision “...with which events are recorded and entered” (Wagner and Hogan 1996).
- Completeness is defined as being all-inclusive. Completeness is achieved when nothing relevant is excluded. Completeness means that “...all necessary values are included” (Wand and Wang 1996).
- Consistency is generally defined as

reliability. Something is consistent when the same result is produced for all subsequent iterations (Wand and Wang 1996).

- Timeliness is believed to include two dimensions. On one hand, timeliness is defined as being up-to-date. On the other hand, timeliness may be defined as the actual number of seconds, minutes, hours, etc. required for a task to be completed (Wand and Wang 1996).

The MIWASS team believed that users would find that the MIWASS dental and veterinary modules on the PDA platform produce high quality results in terms of data input and usefulness. This follows the findings of previous studies indicating that the use of PDAs in the medical setting leads to better quality as opposed to other methods, such as paper-based techniques (VanDenKerkhof, et al. 2003; Menke, et al. 2001; Bates, et al. 2003).

VanDenKerkhof conducted a study comparing the use of PDAs with paper-based methods for the purpose of gathering patient data, determining that PDA quality was superior to paper-based methods regarding timeliness. Physicians were able to record data at the point of care faster with a PDA than with paper (VanDenKerkhof, et al. 2003). Time savings can be crucial in healthcare because physicians sometimes cannot see patients due to time constraints. Thus, a PDA could result in improved patient care.

In a study comparing PDAs with paper-based methods in a pediatric intensive care unit, Menke found the quality of data completeness was better when PDAs were employed, as compared with paper. Menke reasoned that the use of PDAs resulted in higher-quality documentation than paper. "In all areas, nursing documentation was superior during the CDS [PDA] time period both in content and legibility" (Menke, et al. 2001).

In a position paper comparing electronic medical records to paper-based records, Bates suggested that primary care facilities should abandon paper-based techniques and adopt PDAs for the purpose of record-keeping. Bates argued that the following are disadvantages of paper-based records: "...availability to only one person at a time; frequent illegibility; inability to be accessed remotely or at the time and place needed; growing so thick as to be unwieldy; low utility and large overhead as vehicles to evaluate quality; and segmentation with multiple volumes and multiple storage sites." Bates insisted

that the aforementioned pitfalls would be circumvented by using PDAs (Bates, et al. 2003).

2.1 EFFICIENCY

The MIWASS dental and veterinary modules were measured for efficiency using the above-mentioned survey.

The MIWASS team predicted that users would consider the MIWASS dental and veterinary modules efficient for their purposes. Other studies suggest that physicians and nurses rated PDAs as highly efficient in their respective settings (Poissant, et al. 2005; Rothschild, et al. 2002).

In a study on the impact of electronic health records on the efficiency of nurses' documentation time, Poissant discovered that electronic health record time efficiency is superior to other methods. "The use of bedside terminals and central station desktops saved nurses, respectively, 24.5% and 23.5% of their overall time spent documenting during a shift" (2005). Time saved could lead to improved patient care.

Rothschild conducted a study on the efficiency of handheld devices for use as drug reference guides by clinicians. He reported that clinicians considered the handheld device more efficient in regard to time than other methods. "Clinicians in our survey most frequently needed one to five minutes to find drug information when using traditional search methods, but they reported needing less than 20 seconds when using (handheld devices) Rx" (Rothschild, et al. 2002).

USABILITY

There is widespread agreement in the literature that usability is multidimensional. Shackel defines usability with the following attributes: effectiveness, learnability, flexibility, and attitude. Effectiveness is realized when a task takes a shorter time to complete, and the finished task contains fewer errors. Learnability refers to the length of time it takes to learn a task and the amount of information that is retained. Flexibility is defined as the degree to which users are able to adjust to changes in tasks. Attitude is expressed as a user's overall subjective likeability of performing a task (Folmer and Bosch 2004).

The MIWASS dental and veterinary modules were tested for usability by the user group. personnel evaluated the field testing results using the Task-Technology-Fit (TTF) model. The TTF model is a popular model for obtaining user evaluations of information technologies, such as the MIWASS

dental and veterinary modules. The TTF model is illustrated in Figure 1.

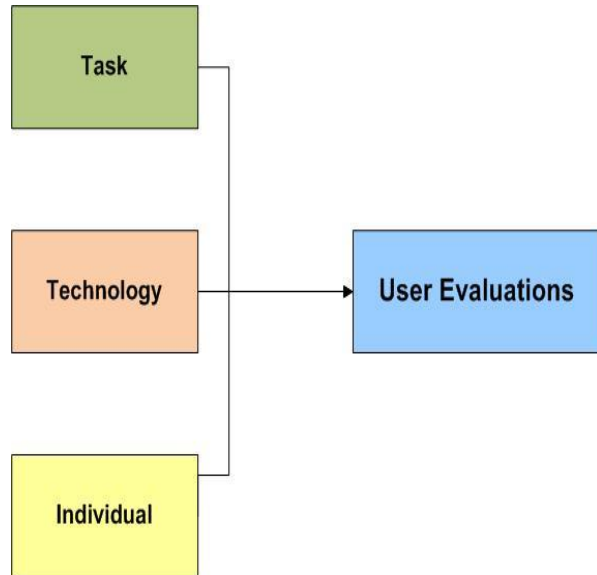


Figure 1. Task-Technology-Fit Model

The central premise for the TTF model is that "users will give evaluations based on the extent to which systems meet their needs and abilities." For the purpose of our study, we define user evaluations as user perceptions of the fit of the MIWASS dental and veterinary modules, based on their personal task needs (Goodhue 1995).

Goodhue and Thompson shed light on the linkages between information systems and individual performance. Their research proposed a "comprehensive theoretical model that incorporates valuable insights from two complementary streams of research." These two research streams are user attitudes as predictors of utilization and TTF as a predictor of performance. The research also empirically tested the core of the model. This model asserts that information technology impacts individual performance in two ways: the technology must not only be utilized, but it must also be a good fit with the task it supports. This model is called the Technology-to-Performance Chain (TPC). The TPC model was used to evaluate whether information technology and services could be used to determine that user needs were being adequately met (Goodhue and Thompson 1995).

Goodhue reported on the development and measurement validity of a diagnostic tool used to evaluate overall information systems and services in an organization. This instrument is based on TTF

theory and postulates that positive user evaluations correspond to information technology functionality and task requirements. Goodhue further suggests that users require different information system functionalities for that task. The TTF instrument is used to "measure the degree to which an organization's information systems and services meet the information needs of its managers." Validity of this instrument was tested using a sample of 357 users in ten companies. The instrument proved to be an attractive option for both practitioners and researchers attempting to measure organizational information system effectiveness (Goodhue 1998).

ANALYSIS TECHNIQUE

We individually analyzed the responses to all 16 closed-ended items on the user surveys. We calculated and reported descriptive statistics for each survey item, including response frequencies, means, and standard deviations. In addition to descriptive statistics, we assessed the reliability of responses to the various scales using Cronbach's α . Finally, to test Intent to Use as hypothesized by the TTF, a multivariable regression analysis was conducted.

STUDY DESIGN

The MIWASS usability testing consisted of five phases:

1. **Training.** Our personnel provided training materials to the usability testing participants. During this time, personnel were available for hands-on training and technical assistance.
2. **Evaluation.** The participants were asked to operate the PDA software application, completing four test scripts.
3. **Survey.** We administered a survey to the user group via paper forms.
4. **Analysis.** We analyzed the results of the user survey.
5. **Refinement.** Based on the results of the survey analysis, we decided that we might refine the MIWASS software application.
6. **We contacted a pool of 73 conference attendees and invited them to participate in the study. Seventy-two of the 73 subjects agreed to participate in the study and signed consent forms (a response rate of 98.63%). No significant biases were apparent.**

TRAINING

During the first phase of usability testing, our personnel gave the 72 users sufficient training to enable them to operate and evaluate the MIWASS (36) dental and (36) veterinary modules. The training program included instruction, handouts, and hands-on training with the actual modules on handheld devices distributed to the users. A customized training application was installed on each device, and the instructors guided the users in working with it to learn how to operate the equipment and modules.

The instructors ensured that all users were thoroughly familiar with the equipment, modules, and objectives of the study before they participated in the evaluation. The instructors taught all users how to operate the device and module controls, enabling them to follow the steps of operation from start-up to shut-down. The users also learned the steps they would be asked to follow during the evaluation, including entry of data into the modules according to scenarios developed by us. When training concluded, the users were able to switch on the devices, open the modules, enter data according to scenario test scripts, print the form associated with the scenario, close the modules, and switch off the devices.

EVALUATION

After successful completion of training, our personnel asked the participants at the conference usability testing site to complete scripts using the MIWASS software modules installed on handheld devices. Users completed four scripts relating to either the dental module or the veterinary module. The scripts guided the users through the process of completing dental and veterinary forms. During this evaluation period, we used code embedded in the software to capture date/time metrics regarding the length of time it took for a user to enter data associated with each form.

Our personnel were familiar with the modules and scripts and we were on hand to provide support and answer questions. After a user entered the scenarios detailed in the scripts, he or she was able to print the applicable forms using wireless printers supplied by us. At the conclusion of the evaluation, the users were instructed on how to complete the user evaluation survey.

SURVEY

After the users completed the evaluation, personnel distributed a paper form survey to them. A six-point Likert scale was used for measurement purposes.

The Likert scale ranged from “strongly disagree” to “strongly agree” for all 16 of the closed-ended questions included in the survey.

ANALYSIS

Upon conclusion of field testing, personnel collected the devices installed with the MIWASS modules. We reviewed, analyzed, and reported the date/time metrics that were collected from embedded code during field testing. Our personnel calculated and reported descriptive statistics such as the mean and standard deviation for each script.

Data screening was carried out on the set of 72 responses, and the final data set of all 72 usable responses was used in the study (36 dental and 36 veterinary). The respondents completed the entire survey questionnaire.

RESULTS

The results of our analysis revolve around two approaches. One approach features technical data gathered on the duration of each encounter type, for animal exam and vaccinations, dental exam, and dental health record exam forms, and the percent accurate for these encounters.

TECHNICAL DATA ANALYSIS

Thirty-seven subjects participated in the veterinary application testing. The mean duration was 310.49 seconds (5 minutes, 10.49 seconds), with a standard deviation of 97.28 seconds (1 minute, 37.28 seconds). The animal exam and vaccinations ranged from a minimum of 160 seconds (2 minutes, 40 seconds) to a maximum of 640 seconds (10 minutes, 40 seconds).

Thirty-seven participants also participated in the dental application testing. The mean duration was 224.21 seconds (3 minutes, 44.21 seconds). The dental exam and dental health record exam ranged from 46 seconds to 534 seconds (8 minutes, 54 seconds), with a standard deviation of 128.09 seconds (2 minutes, 8.09 seconds).

A pair-wise comparison of means shows that a significant difference exists between the duration of dental and veterinary application times to complete the exercises ($p=.005$).

User errors were most likely to occur when performing free text entries as opposed to check box/drop-down menu entries. In fact, 96.3% of all errors resulted from free text entries. Therefore, we

can tentatively conclude that users had difficulty with the PDA's on-screen keyboard. It should be noted that the on-screen keyboard utilizes the maximum possible screen space on the PDAs used in this study.

PERCEPTUAL DATA ANALYSIS

The second approach of our analysis is behavioral and involves the perceptions of the survey respondents to the applications. Data for each question was entered into a Microsoft Access database. Initial summary statistics, including frequencies, item means, and item standard deviations were then calculated. On a scale of 1.0 to 6.0, with 1.0 meaning the modules were difficult to understand or use and 6.0 meaning they were easily understood and used, the average item response was 5.17, with an average standard deviation of 0.85 (95%CI = 5.17 ± 0.20). Because negative responses ("Strongly Disagree," "Disagree") were assigned lower scores and positive responses were assigned higher scores, this implies that, on average, users tended to agree that the system was easy to use, useful, and fit their technological needs.

Next, the reliability of each scale measured by the MIWASS user evaluation instrument was assessed using Cronbach's α , which is a measure of internal consistency. Both the Task Technology Fit model and the Technology Acceptance Model attempt to measure constructs—namely, ease of use, usefulness, performance impact, and technological fit. Cronbach's α measures the degree to which a set of items consistently measures a unidimensional construct such as these, based on the inter-item and item-to-total correlations. Generally, an α of 0.70 or higher is considered acceptable.

With the exception of the TAM Ease of Use scale, all scale reliabilities were in the high 0.80s and low 0.90s, indicating high internal consistency. With regard to the TAM Ease of Use scale, each item was analyzed to see the effect on Cronbach's α of removing it. It appears that Item 13, which was reverse coded, had very low item-to-total correlation. This is common for reverse-coded items, especially at the end of a survey. Inattentive respondents will answer in the affirmative (i.e., Strongly Agree or Agree), when their true intention is to answer Strongly Disagree or Disagree.

Multiple-regression was performed to assess the contribution of perceived usefulness and perceived ease of use on overall intent to use. The results indicate that overall user evaluations of the IT application use are determined by the TTF model

item's ease of use, usefulness, performance impact, and technological fit. The F value was 16.991, and the model was significant at a $p < 0.001$ level of significance. R^2 for the model was 0.809, which indicates that model independent variables explain about 81% of variance in the dependent variable.

In addition to the model summary information, coefficients for each of the independent variables (items not measuring Intent to Use), along with t-test, standard error, and significance were determined. Questions 1, 4, 5, 7, 8, 11, 14, and 16 were significant at the 0.05 level of significance.

CONCLUSIONS

The contribution of our research can be summarized as follows: we developed an instrument to measure the usefulness of the MIWASS dental and veterinary modules in healthcare, and we successfully tested the TTF model on those modules. The TTF model is a continuously evolving model for measuring user evaluations of information systems. Though earlier studies (Goodhue and Thompson 1995; Goodhue 1995) focused on developing a multi-dimensional construct for user evaluations of TTF, very little information was provided on how to measure technology, task, and individual characteristics constructs. We used the TTF model to develop an instrument for obtaining user evaluations of the MIWASS dental and veterinary modules. We have shown that the major factors that impacted the overall use of the modules were that they were easy to use, useful, and fit users' technological needs. Our results indicate that TTF, along with individual characteristics, has an impact on user evaluation of healthcare IT.

Survey participants were quickly and successfully trained to use the MIWASS dental and veterinary modules for and were satisfied the modules performed most of their data collection functions very well. Participants also indicated the modules could be useful tools in collecting and disseminating data and had features that would allow users to obtain, evaluate, and present information more efficiently than previous methods. Overall, participants indicated the MIWASS dental and veterinary modules had significant potential digital data collection utility. Based on these results, we recommend that the MIWASS dental and veterinary modules be introduced to the workforce to enable the digitization of dental and veterinary records.

REFERENCES

Will be furnished upon request.