Tangerine™ - Electronic Data Collection Tool for Early Reading and Math Assessments

January 2012 – Kenya Field Trial Report: SUMMARY

Background

Tangerine™ for electronic data collection

Since 2006, early grade reading assessments (EGRA[1]) have been used to rigorously diagnose national and system-level gaps in reading competencies among students in over 50 countries and 70 languages. Results have informed education policy reform and improvement of teacher training around the world, contributing to a reading revolution in international development. While the traditional, paper-based method of recording student performance provides high quality, informative data for understanding levels of skill acquisition, the costs of measuring large samples (3,500 is the average size for national baseline samples[2]) two or three times during a project's implementation is significant, and an estimated 5% to 15% of measurement costs can be directly related to costs of the paper-based administration (logistics, copies, supplies, data entry, etc.[3]). The largest share of these paper-related costs is the preparation of the data entry system and transcription from paper forms into the system for data analysis. This process is also prone to human error, compromising the quality of the data and delaying its availability for analysis and decision-making.

In 2009, Tangerine electronic data collection software was conceived of as a solution to address some of the challenges encountered in paper-based administration, particularly for large, labor-intensive national samples. RTI developed the software over the period of September 2010 to September 2011 as an open-source data collection system for mobile devices; the system allows assessors to record student responses electronically rather than on paper but otherwise does not affect the traditional administration process of EGRA or Early Grade Mathematics Assessment (EGMA). As a browser-based (but offline) survey using cutting-edge open-source technologies optimized for touch screen tablets, Tangerine enables compatibility across a wide range of mobile devices, languages, and writing scripts. A critical feature is that it enables data synchronization from various assessors, working simultaneously offline on different devices, with a central database, over a variety of connection types.

Some of the more general advantages to an electronic approach to survey data collection include the following:

- Rapid availability of assessment data

[1] EGRA and EGMA are timed, oral assessments of early learning skills; the assessments are administered by a trained assessor who provides the child with a set of stimuli (i.e., reading passages or math tasks) and records whether the child can read or calculate correctly across the short tasks, as demonstrated by responding out loud to a set of stimuli.

[2] See EGRA Tracker, at www.eddataglobal.org.

[3] Source: RTI analysis of past budget and modeling of estimated costs of using electronic approach.

- Improved quality and consistency of data and fewer measurement errors due to missing fields, data transcription errors, invalid data types or formats, illegible or incomprehensible data, etc.
- Reduced amount of paper and supplies, as well as more manageable logistics required to prepare and manage the paper assessments (photocopying, sorting, stapling, packaging, etc.)
- Possibility of accessing additional information and media within the portable device, i.e., audio files, video tutorials, instruction manuals, global positioning system (GPS), etc.

These advantages to electronic survey data collection have long been leveraged by health and government sectors in high- to middle-income countries and, more recently, also in lower income countries; these sectors employ electronic systems for medical records, census data, household surveys, and more (RTI International[4]; Kleinman, et al.[5]; Haller, et al.[6]). For that reason, many commercial and open-source electronic survey data collection tools exist on the market; however, the timed, oral nature of the early grade reading and mathematics assessments required the development of a new tool.

About the Kenya field trial

Internal and external alpha testing of the software was completed in October 2011 to ensure basic functionality, but it was important to conduct more thorough field testing before using the software for large-scale data collection in case contextual factors, such as user expectations, environmental conditions, or scale of use (i.e., number of consecutive tests conducted and overall size of the database), would uncover any unanticipated requirements that would necessitate modification of the design. In January 2012, the Primary Math and Reading (PRIMR) Program, a 3-year US Agency for International Development-funded EdData II project, implemented by RTI International, provided a site for a small-scale data collection trial using Tangerine. This field trial was intended for a first-level analysis of hardware and software performance, usability, and the overall feasibility of the approach based on actual conditions of administration. It involved a small sample in an urban area of Kenya consisting of 200 children in 10 schools, each completing three assessments (English EGRA, Kiswahili EGRA, EGMA), for a total of 600 assessments and more than 176,000 individual data points captured. Six enumerators visited schools in two teams of three enumerators each—one school per day, testing 10 students from first grade and 10 from second grade.

The key research questions that the field trial aimed to answer were focused on the following:

- Functionality of hardware[7] and software: Does the electronic approach allow for complete and accurate recording of student responses? Does the electronic approach offer a secure method of backing up and transferring data for analysis? Is the hardware appropriate for typical assessment conditions?
- Usability: Have we achieved a *user-centered design,* from the perspective of typical end users, for Tangerine? Does the user interface of the selected hardware model create any barrier to use?

---

4 RTI International. 2004. Handheld Devices: Field Data Collection and Transmission Techniques.

5 Kleinman, L., Leidy, N. K., Crawley, J., Bonomi, A., and Schoenfeld, P. (2001). A comparative trial of paper-and-pencil versus computer administration of the quality of life in reflux and dyspepsia (QOLRAD) Questionnaire. *Medical Care 2001; 39*:181-189. © Lippincott, Williams and Wilkins.

6 Haller, G., Haller, D., Courvoisier, D., Lovis, C. (2009). Handheld vs. laptop computers for electronic data collection in clinical research: A crossover randomized trial. *Journal of the American Medical Informatics Association, 16*(5) Sep/Oct.

7 For this field trial, we chose the Amazon Kindle Fire. Designed as an eReader with multimedia capability, it has a web browser and the possibility to add third-party Android applications in addition to reading e-Books or other digital documents (i.e., pdf or Word documents).

- Comparison of electronic vs. paper assessments: To what extent is electronic administration as reliable as the paper administration? What other added value can be achieved through the use of the electronic mode?

Findings from the field assessment

Functionality

Overall, the Tangerine software performed as it had been designed to: enumerators were able to start the system, start new assessments, call up preloaded data (school names with district, region, and code), create unique student identification numbers, enter performance information on all types of subtests, record answers to the basic student demographic questionnaire, save tests, view cumulative tests collected over the course of the day, and export (synchronize) data to a central database. The system prevented errors common to paper administration such as missing data or incorrect formats. Some issues were also identified where software enhancements would improve functionality (e.g., the frequency with which data is saved locally and options for syncing remotely). The selected hardware, the Kindle Fire, also proved to be easy to adopt by enumerators, responsive, and durable. The battery life was sufficient for a full day's worth of data collection, or at least 25 individual assessments.

Usability

Murphy et al.[8] notes that good user design is critical; when cognitive demand associated with the user interface is more reasonable, users are better able to focus on their tasks. The nature of EGRA/EGMA is such that assessors must be intensely focused on the child's responses, following along at the pace of the child—there is no room for delay or struggling in any way with the technology interface. In this case, there were two user interfaces to contend with: the Kindle interface and the Tangerine interface. The latter runs within the Kindle Silk browser and has its own system for advancing from screen to screen, entering data through radio buttons, check boxes, buttons, or text entry fields.

As a framework for assessing usability, we used the following criteria, cited in Murphy:

- ***Learnability***: is it easy to learn?
- ***Efficiency***: does it allow the user to accomplish the task quickly and effectively?
- ***Memorability***: is it easy to remember how to use it after a long period of non-use?
- ***Error rate***: does it have a low incidence of errors, and is it easy to recover from them?
- ***Satisfaction***: is it pleasant to use?

We found that for both the Kindle interface and the Tangerine interface, **usability** requires becoming comfortable with the use of the touch screen and finding the most effective and efficient "touch." The assessors mostly preferred using the stylus, although this depended on the type of input that the system required; for example, the radio buttons seemed easier to activate with the finger rather than the stylus, but the stylus was by far easier and more accurate for entering text from the virtual keyboard.

Related to ***learnability***, the assessors generally had no trouble learning, understanding, and adapting to the touch-screen logic or the interface of the device, including how to scroll up and down or activate menus

[8] Murphy, E., Marquis, K., Hoffman III, R., Saner, L., Tedesco, H., and Harris, C. (no date). *Improving Electronic Data Collection and Dissemination Through Usability Testing.* US Census Bureau and Computer Technology Associates.

and links with a touch of the stylus. The user group in Kenya had previous experience with paper-based EGRA/EGMA surveys, web-based surveys and touch screen smartphones. The Tangerine interface was also very easy to learn because it closely matched the design of the paper test.

There were both positive and negative tradeoffs in ***efficiency*** of administration with the electronic system. Increased efficiency was achieved by features such as automatic population of the school list (school name, region, district, school code) using a rapid input technique. Also, not having to deal with an 8-to16-page instrument, plus stopwatch and other manipulatives, saved time and created a more child-focused environment. A few issues related to the Kindle Silk browser and the way Tangerine functions caused some loss of efficiency for the new user, and these findings allowed us to enhance the design of the system to avoid these issues. For example, when running a website from a local host, which is the case with the Tangerine system, the browser appends "www." and ".com" to the local URL, so several steps were required to ensure that this did not prevent the application from loading when starting a new assessment. Additionally, through this field trial we were able to determine the compatibility of the dataset created by Tangerine with other standard EGRA/EGMA databases and the level of effort required to prepare the dataset for use. While some data cleaning still has to be done, the electronic system itself does not require any greater level of effort; rather the same process is applied. On the other hand, the data collected electronically is already much "cleaner" than data collected on paper because there are fewer transcription errors (i.e., how a school name is spelled) and the data is much more consistent (i.e., date formats). By eliminating the data entry phase (transcription from paper to computer records), the project eliminates an additional source of human error and saves a significant amount of time and resources.[9]

The short pilot could not answer the question of ***memorability*** immediately, since we were working according to a short data collection schedule and the longest period of inactivity was over the weekend. However, from this experience, it seems that the only difficult parts of the system to remember were the steps for activating Tangerine, which at the time of this field trial required several steps. Development of a combined single-APK file installation package (an Android Package installation file, completed in March 2012) has eliminated this difficulty.

The ***error rate***, defined as the system not behaving as it is expected to, was very low; however, the errors that did occur were important and allowed us to improve the design of the application. For example, some administration errors were encountered as a result of the touch screen interface (e.g., accidentally touching an input button while handling the device). These errors were mostly unrelated to the functional design of Tangerine, but rather were relevant to the visual design and where buttons should be placed on the screen to minimize accidental interference when scrolling or using other native features of the hardware. Additionally, some data loss was experienced due to system "freezes" caused by processing all data from the forms once, at the end of the test (this one-time processing was used in the field trial version of the system). The system freeze problem has since been eliminated by processing the forms after each subtest.

All of the enumerators expressed a high degree of ***satisfaction*** with the approach, compared to paper-and-pencil administration. One reason for this is that there is a reduced number of items to manage—for paper administration: clipboard, paper and page turning, stopwatch, student stimuli, and pencil versus for electronic administration: the Kindle, stylus (optional), and student stimuli—and therefore fewer possibilities for loss, damage, or misuse of any one of the items during the administration process.

[9] In the present case, the time required to enter the data from the number of assessments undertaken would have been 53 hours, or 6.5 days. By using electronic collection, all 53 hours of data entry were eliminated.

Another reason is that the small, handheld, yet firm input surface is much better in the conditions in which the tests usually take place—on benches with no flat surfaces or tables available for placing the items. Instead, the enumerators can hold the devices in a number of different positions while maintaining confidentiality from the student using the cover of the case.

Supplies required for electronic administration (left) for paper administration in one school (center) and for entire data collection. Photos: RTI/S. Pouezevara.

Comparative value of electronic versus paper

During this first major trial of electronic data collection with this software and hardware combination,[10] we were able to identify areas where the application needed to be optimized for accuracy and efficiency of data collection. While some tests were lost due to system failures or administration errors as mentioned above, overall data quality during administration was similar to what is achieved with paper administration, where on average, 2% of tests administered are not used for analysis because of errors or omissions[11]. On the other hand, with the electronic system, data loss for reasons such as technology failure were immediately evident, allowing the data collection team to note the error and conduct the assessment with another, similarly randomly selected child to maintain the required number of observations. Similarly, with the electronic approach, assessment administrators could keep track of fieldwork progress on a daily basis. During the field trial, data from the Kindles were synchronized on a daily basis using a combination of a wireless hotspot and a 3G USB dongle. This allowed us to verify data completeness at the end of every day. Some authors suggest, according to Haller et al. (2009),[12] that the use of handhelds could negatively impact data quality compared to the use of other types of electronic devices (i.e., desktop or laptop computers) because of the small screen size or input format. While we did experience some such issues, e.g., accidental screen taps during routine handling, we expect that in future implementations the longer enumerator training period and inter-rater reliability procedures, combined with feature enhancements identified by this pilot, will minimize their occurrence.

Conclusions

Comparison of electronic and paper approaches was not the primary goal of this pilot; rather, we were testing the functionality of a beta version of the electronic data collection software with several known issues (incomplete features, bugs). However, the trial allowed us to make some broad conclusions that confirm the feasibility and merits of the electronic approach over the paper equivalent: for example, rapid availability of results, error-checking during administration, and reduced time and effort required for

[10] RTI has also used an electronic version of EGRA designed by a Yemeni subcontractor for use on the iPad, in Arabic.
[11] RTI International, 2012 internal analysis of datasets.
[12] Haller, G. et al. (2009). See full citation in footnote 6.

preparation of fieldwork logistics (photocopying, stapling, etc.). We were also able to test the feasibility of using a larger, more conventional laptop (netbook) as the data entry hardware. In this case, we found that the touchpad input was not accurate enough or responsive enough to keep pace with the child's reading; yet in most typical EGRA situations, it is not practical to use a more responsive (and familiar to most enumerators) external mouse for input because there is rarely a sufficiently large or smooth table to sit at. Additionally, we observed much more disruption in the classrooms and distraction by the children when the netbooks were used compared to when the Kindles were used. Therefore, we consider the tablet to be more appropriate for large-scale assessments.

The field trial also allowed us to review the entire workflow, from both an administrative perspective and an enumerator perspective, and ensure that all functionality is optimized, from test development (adaptation to the electronic format) to pilot testing, making revisions and updates, assigning users, collecting data, sending/merging data, cleaning, and analysis. Several areas were identified for optimization, most notably the packaging and installation of the system, including updates.

In terms of costs, there are tradeoffs in the initial purchase of expensive hardware (the Kindle Fire is the most reasonably priced for the quality of the hardware, at $199) and anticipated recurrent costs. From past analysis of EGRA budgets, the costs of paper-related inputs can reach up to 15% of the total assessment budget. A rough comparison of paper versus electronic was made using information from the Kenya trial, comparing the unit costs of implementing the national assessment on paper (sample size of approximately 4,400 children) versus the projected costs of using the electronic data entry system. In this case, the costs were nearly equivalent. (In other words, the costs of the initial hardware purchase would have been completely offset by reduced inputs related to the paper administration, as required for that sample size). This substantiates previous estimates by RTI that concluded that the electronic approach may not be cost effective for a single national baseline (depending on size and local unit costs for things such as labor and supplies), but when a project expects to do baseline, mid-term, and final evaluations, the initial investment is paid off quickly. Additionally, the tablets can serve many other functions related to improving educational administration and pedagogy when not in use for the field surveys. Under PRIMR, for example, they will be loaded with various electronic teaching and learning materials that instructional coaches can use during school support visits.

Further small- and large-scale field implementations are being planned and will be used to test recent enhancements and constantly improve the quality and usability of the application. Although Tangerine was developed initially by RTI International through internal investment funds, as an open-source and non-commercial tool, it is improved each time it is adopted and implemented, and these improvements continually benefit subsequent implementations as well. Therefore, just as EGRA was optimized over years of use in various contexts and through collaboration among various organizations and donors, it is expected that the electronic approach will also require several field implementations to enhance reliability address challenges, and improve its role in implementation.

For comments or further information, please contact Carmen Strigel at: cstrigel@rti.org.