Baccalauréat Practice Tests in Cameroon: The Impact of SMS-Based Exam Preparation

Anthony Poon
Cornell University
New York, New York
anthop@infosci.cornell.edu

Sarah Giroux
Cornell University
Ithaca, New York
sh104@cornell.edu

Parfait Eloundou-Enyegue
Cornell University
Ithaca, New York
pme7@cornell.edu

François Guimbretière
Cornell University
Ithaca, New York
fvg3@cornell.edu

Nicola Dell
Cornell Tech
New York, New York
nixdell@cornell.edu

ABSTRACT

In Cameroon, no exam is more important than the baccalauréat, which serves as the gateway to adulthood and professional and higher-education aspirations. We build on existing literature on the benefits of practice tests to evaluate how SMS-based quizzes could help students in their final year of secondary school prepare for the baccalauréat. Students in 23 schools across Yaoundé, the capital of Cameroon, received multiple-choice questions four times per week for nine weeks leading up to the exam date. We examined baccalauréat pass rates and student feedback through 11 focus groups. We find that students used our practice tests for formative assessment and as a prompt for recall and review of study material. SMS-based quizzes were not isolated artifacts but were shared in the students' networks, built a distributed learning environment, and used as a focus of collaborative study sessions.

CCS CONCEPTS

• **Human-centered computing** → *Mobile phones*; **Empirical studies in HCI**; *Empirical studies in collaborative and social computing*; • **Applied computing** → **Computer-assisted instruction**; *Psychology*; Collaborative learning.

KEYWORDS

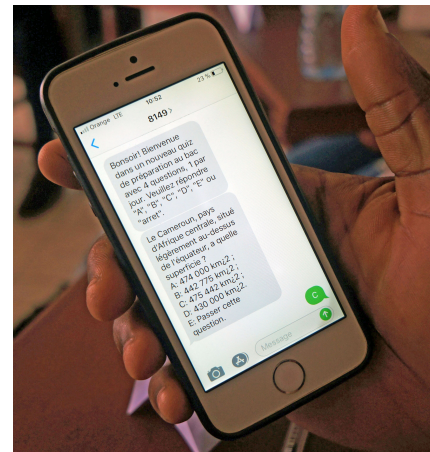education, practice tests, standardized tests, testing effect, field study, SMS, ICTD, HCI4D

Figure 1: The practice tests on a student's phone. This shows an introduction message with instructions sent at the beginning of each week, a multiple-choice question, and a student reply with their answer choice. A feedback message would be sent the following day providing the correct answer.

1 INTRODUCTION

In Cameroon, a sub-Saharan nation in west-central Africa, secondary school students spend years preparing for what is likely the most important final exam of their lives, the baccalauréat or "bac." This qualifying exam serves to mark the successful completion of secondary school and entry into adult life. The diploma conferred upon passing is important for future employment prospects or admission into higher education. Students and their families spend significant effort and expenses towards this exam, including tutors and paid study groups. Despite this, pass rates for the bac have historically been low, with only 60.19% of students passing in 2019 [6]. Failing the bac can result in repeated years of schooling and additional financial burden on students and their families.

Because of its importance in students' lives, education is a domain of considerable interest for ICTD scholars. Past research includes how ICTs impact the practices of teachers [16, 73] and influence the direction of entire educational systems [5, 71]. Approaches in human-computer interaction for development (HCI4D) have used technology to reconfigure how students learn [43, 58]. This

follows a long thread in HCI scholarship in learner-centered design, which uses educational theory to inform the design of software [69] and hardware environments [44] to support students' learning needs. We build upon this literature by applying cognitive theories around exam preparation to education in the Global South.

The traditional role of tests and exams are to assess student knowledge and assign them grades or scores. However, tests can also have learning benefits of their own, and tests designed for this are called *practice tests*. These are usually smaller quizzes or evaluations, either self-directed or provided by the instructor, with low or no-stakes. Past research in practice tests suggests several ways in which they can enhance learning. Such benefits include strengthening memory pathways for the recall of information [7, 8], early assessment of gaps in knowledge to help focus further efforts in studying [23, 51], and repeated and spread out study [46]. Through these benefits, research has found that practice tests can improve student performance on final exams [11, 23].

Our goal was to examine how SMS-based practice tests could help students study for the bac. Previous work, mostly in Western contexts [15, 19, 38], has described potential benefits of practice tests for student performance on classroom summative exams and grades [23, 48, 76]. In contrast, we address student preparations for a standardized exit exam widely administered in a non-Western country, Cameroon. We examine the impact of our intervention on student study habits and preparations.

We worked with secondary school teachers to write multiple-choice questions for various subjects covered in the bac. We built an automated system to administer practice tests directly to students' personal phones outside of the classroom. We deployed this system to 500 students across 23 public, private, and religious schools in Yaoundé, the capital of Cameroon. We conducted 11 focus groups with 58 students to gather qualitative feedback on how the SMS-based practice tests impacted their studying and what they perceived to be the benefits and drawbacks. We collected bac pass rates for students who received our quizzes and an additional 1424 students in control groups.

Our findings suggest that students in our study experienced many of the benefits described in past literature on practice tests and also some unexpected benefits. Although we were unable to detect a significant difference in bac pass rates between students who received the practice tests and those who did not, we use students' experiences to build on past literature on practice tests. We provide details for potential mechanisms behind the benefits of practice tests and discuss the implications of practice tests in contexts of collaborative study, used outside of the classroom, and for high stakes standardized tests. We also consider the affordances and limitations of using SMS as a medium for delivering practice tests and offer design suggestions for future work in practice testing and educational content in the Global South.

2 RELATED WORK

Our work builds upon research on educational technology in developing contexts and literature on the effects of practice testing. We extend practice testing research into contexts in the Global South.

2.1 Educational Technology in HCI4D

For as long as ICTD has been a field, researchers have designed educational interventions to enhance the classroom experience. Some of the earliest work used multiple mice to enable collaborative learning experiences among school children sharing limited computing resources [58] and authoring tools for intelligent tutoring systems [16]. Researchers have examined the possibility of using cell phones in the classroom to create educational games or distribute and collect home work [27] or DVDs to replace books [29]. Other research has studied using virtual spaces, such as massively open online courses (MOOCs) to extend the reach of educational systems to underserved populations [22, 45].

A significant portion of the educational technology work in the Global South has focused on informal and adult education, primarily in the health domain, such as software for the creation of training materials for low-literacy community health workers (CHWs) [54] or community video for maternal health education [42].

Due to the prevalence of basic cell phone technology, SMS is a common medium in ICTD interventions [34, 55, 59, 78], and this extends to the education domain. SMS has been used to promote HIV/AIDs awareness in Uganda [21] and diabetes awareness in Egypt [1]. Gill et al. used SMS to enhance CHW continuing education in Vietnam [31], and Poon et al. compared SMS and WhatsApp for delivery of educational quizzes in Cameroon [60].

Our work also included an SMS-based intervention that provided quizzes to students on their own cell phones. We extend educational literature in ICTD by using quantitative and qualitative data to understand how such quizzes could affect students' preparations and whether students derived any benefit. We frame the design and evaluation of our intervention in cognitive learning theories, which discuss the role of quizzes in exam preparation.

2.2 Practice Testing for Learning

The use of quizzes to promote rather than assess learning has a long history in education research. In 1917, Gates found that directing grade school students to perform a self-testing recitation allowed them to score better on an exam administered a few hours later over spending the allocated time reading the material [30]. These tests are now called *practice tests* and are low-stakes exercises intended to improve learning outcomes.

Over the years, a large body of literature has grown to show that practice tests positively affect retention and learning outcomes across a broad range of student ages and domains of study [25, 63]. Practice tests are more effective than simply spending the same amount of time on additional studying [50] and include a variety of direct and indirect learning benefits [25, 66]. Recent research has focused on understanding how these benefits are conferred and the best uses and limitations of practice testing.

2.2.1 Benefits of Practice Testing. The most direct benefit of practice testing is the *testing effect*, where requiring students to recall previously learned information improves their performance on later summative assessments [50]. One explanation for this is that testing requires the student to retrieve past memories, and this strengthens or creates new retrieval routes that make it more likely that this information will be remembered and recalled in the future [7, 8].

Practice testing also has several indirect benefits. One is that such quizzes allow students and teachers to identify areas of lack of understanding earlier in the learning process. Students can then allocate time and attention in subsequent study to address those shortcomings [46]. This is called *formative assessment* and can provide students with useful feedback on their performance whether the quiz is self-administered or evaluated by an instructor [23, 51].

Practice tests are also often delayed from the initial learning episode and thus encourage spaced or *distributed practice*, in which practice occurs at regular intervals over a period of time. Distributed practice can improve learning over bunched practice, commonly known as "cramming" [7, 25, 47]. Researchers found that a greater number of more frequent quizzes is more effective than larger, less frequent practice tests [38]. This is because quizzes can prompt students to study on a regular schedule [46], as students are unlikely to optimally distribute their study time otherwise [70].

Other benefits of practice testing include better recall of unpracticed material due to the activation of associated memories within the same [17] or even different domains [12], improving the mental organization of concepts [79] and consolidation of knowledge [24], and reducing stress and anxiety on summative exams [3].

2.2.2   Best Practices for Practice Testing. Research on the limits and best practices of practice testing are mixed. For example, incorrect answers on multiple-choice practice quizzes can incline students to again choose the incorrect answer later [26, 65]. However, providing *feedback* - i.e., the correct answer to practice questions answered incorrectly - can instead result in improved performance [57].

In contrast, some researchers find that incorrect answers are more valuable than correct answers in enhancing learning and suggest that deliberately taking challenging tests would be a more effective learning strategy than avoiding errors [40]. Most literature agrees that feedback is valuable to students [50]. Delaying feedback may also enhance retention over immediate results by allowing the memory of incorrect answers to dissipate or by providing an additional instance of distributed practice [13].

The literature is similarly divided on the efficacy of multiple-choice questions as compared to short answer questions. Unlike short answer questions, the presence of answer choices exposes students to potential incorrect answers [65], and some researchers have found short answer quizzes to be more effective, perhaps due to being cognitively more demanding than multiple-choice [49]. However, other research did not find a difference between multiple-choice and short-answer effects [24, 52], and suggested that the testing effect is still operative when the formats of the practice tests and summative exam do not match [24, 25].

Practice testing is less effective for students who are already high-performers due to a ceiling effect [32, 38] or when academic integrity is not enforced during practice, allowing students to take shortcuts and reducing engagement with the material. This latter case is especially relevant when practice tests are administered outside of the classroom [72] or when the tests are online [20].

2.3   Practice Testing Tools and the Global South

Many educational technology solutions have been built to facilitate practice testing, whether online tools for traditional classrooms [53] or as part of wholly online courses [51]. Whether the technological

The derivative of the function f defined by $f(x) = (2 x^2 + 2x - 1)/(2x - 1)$ is:
A: $f'(x) = (x - 1) / (2x - 1)^2$;
B: $f'(x) = 4x / (2x - 1)^2$;
C: $f'(x) = (4 x^2 - 4x) / (2x - 1)^2$;
D: $f'(x) = (4x(x - 1))/(2x - 1)$;
E: Skip this question.

Figure 2: An example quiz question used in the study translated from French. The correct answer is C.

medium affects the value of these tests is contested. Some have found computerized quizzes encourage a "quiz to learn strategy" that is distracting and less effective than other methods [10], while others found that administering online quizzes can have large and positive impacts on final exams and grades [4, 11].

Technology tools for practice testing also include interactive tasks integrated into connected classroom technologies that provide students and teachers with immediate performance feedback [68]. One implementation common in university lectures are physical audience response devices, popularly known as "clickers," which allows teachers to administer multiple choice quizzes within the flow of a lecture [3, 67]. However, such tools may be out of reach for many resource-constrained classrooms.

2.3.1   Testing Effects in the Global South. It may not be valid to assume that practice testing techniques and the tools designed enable them will be equally useful in the Global South. Our review of the past research in practice tests showed that it has almost wholly been conducted in developed contexts, mostly in the United States [15, 19, 23, 38, 49], and especially the American Midwest [48, 64, 76]. However, pedagogy is a complex and situated process, and teaching practices - such as practice tests - cannot be successfully transplanted without considering the cultural role of education, the typical practices of educators [37], and learning styles in the destination context [36].

Literature on practice tests draw from cognitive and psychological explanations to describe how these tests confer educational benefits [13, 15, 48, 63]. However, cognitive effects are increasingly recognized as not universal, but heavily influenced by cultural and contextual factors [33, 56, 74]. As such, our research question is determine whether a practice test intervention produces these or different effects when deployed to students in Cameroon, a non-Western, lower-middle income, country [28] with a different educational system than in prior practice test literature.

While there is some prior work in developing contexts which addresses participation in quizzes in formal education [60] and the effectiveness of quizzes in informal education [21, 31], to our knowledge, there is no work addressing the testing effect and other benefits of practice tests in subaltern educational systems. We highlight how our SMS-based intervention was designed to produce practice testing benefits, deploy and analyze its impact, and then discuss how it was perceived and utilized by students.

3   DESIGN OF INTERVENTION

Our intervention delivered multiple-choice practice questions to students' personal phones via SMS. We chose to use SMS due to

the high rate of cellphone adoption in Cameroon [2], and because of affordances provided by its push-based and near-synchronous nature [61, 62]. We built an automated system to administer the tests and evaluate student answers. Students could respond toll free by replying with the letter of their answer choice.

Students received four questions per week, one per day, Mondays through Thursdays. Feedback was sent the following day, on Tuesdays through Fridays, and consisted of the letter and short description of the correct answer to the previous question. Finally on Fridays, students also received a summary of their performance for that week as well as their performance for up to three prior weeks. Students were sent all messages regardless of whether or not they responded to any questions.

Design decisions were made in collaboration with local educators and informed by feedback from a previous pilot study. For a more in-depth discussion of the quiz interaction and the technology behind it, we invite the reader to review our prior work [60]. In this paper, we focus on the aspects of the design related to practice testing.

3.1 Embedded Design Intentions

As discussed in Section 2.2, low-stakes quizzes can improve students' academic performance on summative exams by asking students to recall previously learned material, allowing students to assess their mastery of the tested knowledge, and prompting students to study in a distributed and more frequent manner. While there are additional benefits to practice tests, these are the benefits we focused on in the design of our quiz intervention.

3.1.1 Invoking the Testing Effect. We aimed to invoke the testing effect by covering material that would have been previously studied in secondary schools in our context. We worked with local secondary school teachers to write multiple-choice questions for various subjects in the baccalauréat. Subjects covered included chemistry, English and French language, geography, history, literature, mathematics, natural science, philosophy, and physics. Questions had either three or four answer options, and an explicit option to "skip this question" was added to encourage responses and engagement even if a student did not know the answer. Example questions are shown in Figure 1 and Figure 2.

Another way we designed these practice tests for the testing effect was to cater the set of questions to each student taking the quizzes. This was necessary because the baccalauréat had several series with different mixes of subjects of varying weights and distinct diplomas conferred at the end. We asked students which bac series they intended to take and weighted the occurrence of questions to match the corresponding weights of subjects in the reported bac series. This increased the relevance of the quizzes to students in different series and improved chances that students would receive questions related to their own curriculum.

3.1.2 Encouraging Formative Self-Assessment. We included two types of feedback in our quiz intervention to encourage formative self-assessment. Most directly, we grouped questions by subject per week and included performance feedback at the end of the week, describing to students the number of questions answered correctly that week. This allowed students to more easily relate poor performance in a week to weakness in a subject area. We

also provided a record of the students' performance for up to three previous weeks to encourage goal setting and progress tracking over longer periods of time.

Even if students did not answer any quiz questions, we provided answer feedback, a short description of the correct answer, the day after each question was originally asked. This short delay allowed students to retain some context, as the previously received message would be the relevant question and still be visible even on relatively small screens. At the same time, the delay between receiving the question and the answer encouraged spaced practice.

This delay also allowed us to send the same questions to multiple students without some students receiving the answers earlier and sharing them with friends. While we were not concerned with students discussing questions with each other, we did not want them to copy answers without engaging with the questions, reducing the value of practice [72]. By synchronizing answer feedback via SMS, we hoped to avoid students taking shortcuts that would otherwise be possible on quizzes outside of the classroom.

3.1.3 Prompting Distributed Study. Finally, we prompted students to study in a distributed manner by scheduling a message to be sent to a student every weekday. Questions were sent in the evenings to avoid conflicts with daytime activities and match typical hours for studying. Because the quiz intervention would automatically move on to the next question after 24 hours, students were encouraged to answer when they received these questions or their answers would not be counted. This strict timing was enabled by the SMS-based and automated nature of our intervention design.

In addition, our practice tests contained a mix of simpler and more difficult questions. Challenging questions, such as math and physics problems, English translation, and reading comprehension, could not be answered by simple recall. Instead, we structured these questions to require the student to deliberate on the problem or look up required information. By providing challenging questions on a timed basis, we hoped to encourage students to spend time studying on an evenly spaced schedule.

4 FIELD STUDY AND FOCUS GROUPS

We evaluated the impact of SMS-based practice tests via a field experiment with 1924 students in their final year of secondary school in Yaoundé, Cameroon. These quizzes were delivered over a nine-week period from the start of March 2019 to mid-May. We conducted focus groups with students and collected data from student participation rates and bac pass lists.

All procedures were approved by two IRB boards and permitted by the Cameroon Ministry of Scientific Research and Innovation. Additionally, we collaborated with PICHNET, a small research-oriented NGO based in Yaoundé. They were involved throughout the design of the quiz instrument and field study to ensure that both were appropriate and did not harm students.

4.1 Intervention Procedure

4.1.1 Study Sites. Though Cameroon is officially a bilingual French and English-speaking country, most students in the east and south regions attend the francophone education system. Our study sites included 23 francophone schools in Yaoundé that offered terminal year studies. These schools were selected in partnership with the

regional office of the Ministry of Secondary Education. As the office was interested in targeting and addressing student achievement in this project, our sample consisted of mostly schools in the bottom half of bac pass rates in the city.

At least one school was chosen for each of the seven boroughs of Yaounde to give us a broader view of student attitudes and experiences, as neighborhoods within the city were qualitatively distinct. Different neighborhoods were dominated by certain ethnic tribes or socioeconomic classes which may influence schools and the educational aspirations and performance of their students.

Cameroonian schools varied widely by curriculums, resources, and quality of education. Our sample included a mix of public, private, and religiously affiliated schools. Secondary education was not compulsory or free, and both pubic and private schools charged tuition. Public schools tended to have larger student bodies and class sizes than private and religious schools. The average annual tuition for public schools in our sample was about 48,000 Central African Francs (XAF) (about $87 USD) compared to the average of 185,000 XAF ($337 USD) for our private and religious schools. As a national average, education accounted for 9% of household expenditures in 2007 [77]. Our sample included seven public schools, 16 private schools, and one religious school.

4.1.2 Recruitment and Sampling. Staff at our local partner, PICH-NET, helped recruit participants, served as a local point-of-contact for students, and provided logistical support. PICHNET's local contacts also helped us recruit educators to contribute content to our intervention and negotiate access to schools and bac results.

We recruited students in-person with a paper consent document and an Open Data Kit-based recruitment survey through December 2018 to January 2019. With the help of the Ministry of Secondary Education, PICHNET worked with school administrators to set aside some time during normal class hours to perform the recruitment. Field staff gave a short presentation describing their NGO, the purpose of the study, and that participation was optional. Recruited students provided their mobile phone numbers, basic demographic information, and their plans for the baccalauréat. We filtered out students who gave invalid phone numbers or reported that they were not planning to take the bac that academic year. Regardless of consent to participate in the SMS study, students received 4,000 XAF ($7 USD) for their time during recruitment.

In total, we recruited 1924 students. From this sample, we randomly selected two groups of 500 students. The first group received our practice test intervention. The second group received positive messages that simply encouraged students to study. This group acted as a control against the effects of receiving any messages at all. 924 remaining students received no messages.

4.1.3 Practice Test Intervention. Students in the practice test treatment group received one question per day, Mondays through Thursdays, and answer feedback the next day. To improve the relevance of questions for students taking different series of the bac, we subdivided our treatment group to assign different "courses" of practice material. This was based on the bac series that students intended to take, which was self-reported during the initial recruitment.

We built three sub-groups: students who would be taking a series focused on earth sciences (e.g., *baccalauréat scientifique: sciences de la vie et de la terre*), those focused on literature (e.g., *baccalauréat*

littéraire), and those testing in the social sciences (e.g., *baccalauréat économique et social*). Students who reported an "other" form of bac were grouped with the social sciences, as this series was the most balanced in subject weighting. In total, we had 158 students in the earth sciences group, 267 students in the literature group, and the remaining 75 students in the social sciences group.

4.1.4 Positive Messaging to Encourage Additional Study. "Just-in-time" positive motivational messages have been used in many domains to encourage behavior change [35, 39, 75]. In our study, these messages served as an additional control and were intentionally scheduled at the same times and rates as the quiz messages to prompt students to study at those times. Thus, the positive messages were designed to capture some of the benefits of additional study and spaced practice. By comparing the experiences of the practice tests group and positive messaging control, we hoped to determine if there was something about the quiz that would benefit students beyond the effect of encouragement and study reminders.

Students in the positive messaging group received one message per day, Monday through Friday, during the same nine-week intervention period. These messages were adapted from a set written by a local motivational speaker. In general, these messages contained advice and value-laden statements of societal mores interspersed with encouragement to study diligently for the bac.

4.1.5 Intervention Timeline. The intervention period for both practice test and positive message groups began in March 2019. The weekend before the first quiz or positive message, students in both groups received a series of introductory messages describing the purpose of the intervention they were about to receive. We told students that participation was free and any SMS replies would not cost them any airtime. Finally, students were reminded of PICHNET and the recruitment process earlier in the year and given contact details for a staff member to answer any questions that they might have as well as the physical address of the PICHNET offices.

This introduction was intended to build legitimacy and trust in the intervention in hopes of encouraging higher participation rates. Starting the intervention in March, when students were still attending classes, also contributed to this end by allowing students ask their classmates about the messages they received.

The treatment period lasted nine weeks until the end of May 2019. The month of May was a period of independent study with no regularly scheduled classes, and students prepared for the bac outside of the school. Students received at least one message per weekday. Unless a student unsubscribed, a total of 36 quiz questions (or 45 positive messages) were sent during the intervention period. Students received their final message the weekend before the first bac was scheduled to prevent overlap with the exam period. This closing message notified them of the end of the intervention and wished them luck on their upcoming exams.

Students sat for the bac at the end of May 2019. After all bac testing was completed but before results were announced, we conducted focus groups over the course of two weeks at the end of June 2019. Results for the bac were publicly released in mid-July.

Anthony Poon, Sarah Giroux, Parfait Eloundou-Enyegue, François Guimbretière, and Nicola Dell

4.2 Data Collected

In addition to participation rates measured by responses to the practice test intervention and some demographic information from the recruitment survey, we invited students to focus group discussions and collected publicly posted bac results.

4.2.1 Focus Group Discussions.
PICHNET staff recruited students for focus groups from our quiz and positive messaging groups by calling them via their provided phone numbers. We explained the purpose of the focus groups and that participation was optional, and students filled an additional consent form. Focus groups were held at the PICHNET offices which were centrally located within Yaoundé. Students who participated in our focus groups were compensated an additional 4,000 XAF ($7 USD).

Focus group discussions centered on understanding students' experiences in our study and preparing for the bac. We discussed the resources and methods that students used to prepare, how they perceived and interacted with the intervention messages, how it could be improved, and whether they shared the messages with others. We did not recruit students from our control group as we did not expect to get valuable feedback from students about messages which they did not receive. We did not ask questions about students' performance to avoid discomforting participants.

We conducted a total of 11 focus groups with 58 students (33 women and 25 men). 33 students were in the practice test treatment group, while 26 received positive messages. Students came from 12 different schools: 32 students from five public schools, 16 students from six private schools, and 10 from one religiously affiliated school. Five of the focus groups included only students who received quizzes, while two groups included only positive message recipients. The remaining four groups included a mix of students.

All focus groups were conducted in French by local, multilingual moderators, and lasted on average about 1 hour and 30 minutes. We worked with the moderators to translate the questions from English and gauge their appropriateness. The sessions were audio recorded, transcribed, and translated into English for analysis. We used thematic analysis for our focus group data [9]. Codes were generated by reading and re-coding the transcripts multiple times. Examples of low-level codes were: *"liked immediacy of quiz answers"* or *"liked the diversity of questions."* Codes were then grouped into themes, which were reviewed and restructured. Our findings are drawn from our high-level themes.

4.2.2 Baccalauréat Pass Lists.
In July 2019, bac results were publicly posted by testing centers throughout Yaoundé (Figure 3). School-level summaries of results were also typically published in newspapers. These results were *pass lists*, naming all the students who passed the bac from that testing center. If a student passed with honors, that honor was also mentioned in the pass list. PICHNET staff members collected these results by physically visiting testing centers and making photocopies of posted results.

Testing centers for the bac were typically schools, but not all schools offered examination or for all series of the bac. Schools often partnered with each other to ensure that their students had a testing center for their particular exam. We retrieved bac results from 14 testing centers in the city of Yaoundé, focusing on schools either directly in our sample or partnering with schools in our
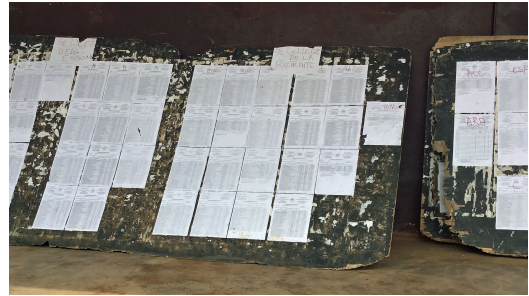
Figure 3: Results for the bac posted outside of the school Lycée Elig-Essono. This location served as a testing center for multiple other schools. Photograph by Valdez Djampop.

sample. Though it was possible there were cases where students took the bac elsewhere, excluding their result from our data set, we believed these to be rare cases that would not systematically skew the accuracy of our analysis.

Examination results were listed by the students' full name. While examinees were assigned a unique registration number, this number did not match up with other identifiers known to us, and thus we could not use it to identify our students. However, because none of our students in our sample reported having the same full name, we were able to use student names to match bac results for students in our sample. We used double-entry by two different staff members to reduce errors in the digitization of bac results.

5 RESULTS

Our results include a quantitative description of participation in our quiz intervention. We discuss students' assessment of the studying benefits of our practice tests and the use of SMS. Finally, we provide an analysis of baccalauréat pass rates.

5.1 Active Participation in Quiz

We define *active participation* as students replying to a quiz question that could be parsed and understood by the automated system managing the practice test. About 228 students, or 45.6% of the students in our quiz treatment group, actively participated at least once during the nine-week intervention period. This is likely a slight underestimation, as there were three days in which an operational issue prevented student responses from being handled and recorded. However, 37 students unsubscribed during the quiz.

Overall, participation was substantially higher than that reported by a similar intervention in Cameroon, which was 36.4% of students [60]. This may be due to greater effort in this study introducing the quiz and the organization behind it and providing avenues for students to determine their legitimacy, changes based on recommendations from that prior work. For example, we added support structures, such as a physical address and manned telephone line, where concerns could directed to improve trust. We made efforts improve student engagement by making quiz content more relevant and challenging and extending the intervention period. Finally, our quiz was structured to send more messages to students even in cases where they did not actively respond to questions.

5.2 Experienced Benefits of Practice Tests

From analysis of our focus group data, we identified three key ways in which students perceived practice tests to help: formative assessment, cued recall of past-learned material, and review of related material.

5.2.1 Formative Assessment Informing Study Strategy. Students used the quizzes to discover gaps and weaknesses in their knowledge. During our focus groups, some students professed to using a strategy that concentrated their efforts on reviewing material with which they were mostly familiar. This allowed them to bolster their scores in these subjects, but meant they were potentially conceding lower scores in other subjects. While not all students described using this strategy, for such students, the formative assessment benefits of a practice test could be eye-opening:

> "There were questions in English, and that's where I really saw that I did not know ... and I realized that I really had gaps in English, and it allowed me to review." [FGD04:Quiz:P4069, Woman]

Students reported surprise at being scored incorrect on questions they thought they answered correctly and were frustrated and discouraged. Such discouragement could reduce future participation, but other students described these cases of incorrect answers as useful opportunities to fix areas of weakness. Students used these questions to discover and learn new knowledge, as described below:

> "It also helped me because I had not read all the literary works, and in philosophy I also discovered certain things that I did not know." [FGD01:Quiz:P4698, Woman]

5.2.2 Testing Effect through Cued Recall. Though not necessarily written by their own teachers, students felt that practice test material was relevant to their past studies, and it cued recall of previous learning experiences. For some questions, students reported being able to answer the question from memory, and in such cases, the practice test exercised students' retrieval routes. In the following quote, one student described how some questions encouraged him to remember older material from a prior school year:

> "[The quiz] allowed me to recall some of the old stuff, even if it was not directly related to the lessons we saw in the terminal class." [FGD04:Quiz:P2251, Man]

5.2.3 Cued Review of Related Material. In some cases, the quiz content was different enough from what students experienced in their own classroom that students never learned the material. Students could not answer such questions from memory. However, a few participants still reported benefits from these questions by being driven to resources such as past notes, classmates, teachers, and the Internet to find the answer. As one participant described:

> "My teachers were my friends especially, so when I came with a question ... we try to debate it, and it becomes the subject of a tutorial session. Because he explains the answer, and he starts to show the source of where it comes from, we remember something else apart from the answer we received by message." [FGD04:Quiz:P2975, Man]

Challenging questions could also support student's learning by pushing them to study and review related classroom materials, as one participant describes below:

> "It pushed me to look things up. Especially when a question came, and I did not have the answer, so I went to look through my notebooks. And the time that I spent looking for the answer, I realized I was already reviewing." [FGD02:Quiz:P3113, Woman]

By review and learning of related materials, students could improve their memory of other knowledge in the same domain as the question prompt. It is potentially through this mechanism that the testing effect can improve student performance on summative exams that inexactly match the curriculum of the practice tests.

5.3 Practice Tests for Collaborative Study

Like the participant who took practice questions to his teachers, several students described how they chose to share their practice tests. Unlike those administered in a controlled classroom, our quizzes were received at home. This enabled students to share them with a wide variety of parties such as students in their cohort, higher-level students, family, and teachers. In this section, we describe some of the reasons that students chose to share their quizzes.

Several students in our focus groups sought help for challenging questions from teachers, elder siblings, and friends already in university. This was particularly true for subjects on which students were less familiar. For example, one participant described taking some questions to her sister who was stronger on certain subjects:

> "I was with my sister. Because we are close, when a message arrived, if I had some difficulties, I showed her the problem. She helped me if she knew the answer." [FGD04:Quiz:P3641, Woman]

Students also shared the messages with peer study groups. This could be partly to work on questions together or to allow friends who did not receive the questions to also benefit. Sharing could occur physically during study meet-ups or by forwarding the messages to others via SMS or other web-based platforms. A few students reported having studying-oriented WhatsApp groups for this purpose. One participant below described turning the quizzes into a game shared among her WhatsApp group:

> "[My classmates] told me that when I receive the messages, I should send them their way so they can also answer. Finally, we turned it into a game; so when I received the messages, I forwarded them, they answered. And even when I was sent the answers, I also forwarded them to check if what they did was right." [FGD02:Quiz:P2387, Woman]

Some students shared their answers with each other to compare notes. This allowed them to debate their respective answers, which may have strengthened recall of the material by comparing their reasoning. In general, students did not consider sharing the questions or seeking help to be cheating due to the preparatory nature of the quizzes and lack of controls against cheating.

5.3.1 Positive Messages Less Likely to Be Shared. Like quizzes, we found that positive messages also prompted distributed practice due to being on the same daily schedule. One student described

how the encouragement he received helped him spend more time studying math, a subject he was not very comfortable with:

> *"At one point I had stopped studying a little bit, but when I started receiving text messages, I went back to work. … I had a very big problem in math because there were a lot of things I did not understand. With these SMS, I fought."* [FGD03:PosMess:P3150, Man]

Students in our focus groups were more likely to describe positive messages as personal and were less likely to report sharing them as compared to practice test participants. 12 of 32 students who received quizzes in our focus groups explicitly described sharing their messages compared to 3 of 26 who received positive messages. 2 practice test students explicitly reported not sharing their quizzes as compared to 6 positive message students. Thus, positive messages, even if they prompted our students to study, did not prompt students to do so in a collaborative manner.

5.4 Affordances and Drawbacks of SMS for Practice Testing

Students also commented on the affordances and limitations specific to the use of SMS in delivering practice tests. Benefits of SMS are driven by the fact that the medium enables push-based interactions, which can be initiated without actions on the student's part. This is more difficult to achieve on both paper and web-based tests that are intended to be taken outside of the classroom. Disadvantages of SMS are primarily caused by the limits of text and high cost of SMS. We describe how students experienced these aspects of the SMS medium with specific relevance to practice testing.

5.4.1 Passive Participation. The push-based nature of our quizzes meant that some students could potentially participate by simply reading the quiz questions and performing a self-assessment when the correct answer was received. Though students were told that SMS replies would be toll-free, a few respondents, like the participant below, read the messages but were afraid of answering due to concerns that they would be charged money to reply or participate:

> *"I still read from time to time … I did not try [to answer]; I was afraid that it would use my credit."* [FGD02:Quiz:P4662, Woman]

Avoiding the cost of replies did not stop students from reading the questions or performing a self-evaluation when they received the answer. We define this as *passive participation*, in contrast to the active participation that is visible to the automated system. The nature of the SMS medium and the fact that entire quiz content was sent via SMS enabled this passive mode of participation.

5.4.2 Reliability of Feedback. Feedback was also a pushed message, which increased students' perceptions of its reliability. Because answer feedback was always sent the day after the question was asked, several focus group respondents described that they appreciated being able to depend on receiving the correct answer the following day and were encouraged to participate because of this regularity. One student described the rapidity of answer feedback:

> *"The next day before asking a new question, there was already the answer."* [FGD08:Quiz:P3088, Man]

We were concerned that students would lose the relevance of the correct answer due to this one-day delay, especially since it is relatively common practice in our context for students to quickly delete older messages to free up memory space. However, this turned out to be a non-issue, and none of our participants described problems relating answer feedback back to the original question.

5.4.3 Limitations of Plain Text Interfaces. SMS's primary limitation is its text-based nature. Though pictures can be supported through MMS, this comes with additional airtime costs for both the quiz administrator and the student. Questions in some subjects, such as math, physics, and chemistry, were poorly suited to SMS due to this limitation. Even questions which could be stated in text were harder to understand due to limits on text formatting and special characters. For example, the question in Figure 2 would be more intelligible with super-scripting instead of using the carat character for power operators. One participant pointed out this limitation on basic phones, and suggested a multimedia enabled digital platform would better suited for this content:

> *"For example, in mathematics, there are equations that cannot appear in some ordinary phones like mine, so you have trouble understanding what the message means. It's important to keep the format in mind, so why not offer a digital platform...."* [FGD06:Quiz:P0659, Man]

5.4.4 High Costs and Short Lengths. SMS also is costly on a perbyte basis, and not all students have smartphones with large screens. These both meant that it was difficult to send large amounts of text at once. Thus, when we sent feedback, we were unable to include a large explanation of how the correct answer was reached. This can add to the opacity of answers and increase student frustration. One participant described the need for more in-depth explanations:

> *"When you send the answer, I tell myself that there are some questions that will need to be explained. It is not enough to just send the answer."* [FGD06:Quiz:P4684, Woman]

5.5 Baccalauréat Pass Rates

We collected bac pass rates in July. Overall, 657 students of our 1924 recruited students passed the bac, for a total pass rate of 34.15%. Of these, the vast majority of students passed without any special honors. 48 students passed with *assez bien* or "pretty good," five with *bien* or "good," and none with any other honor. The pass rate of students in our study was much lower than the national average of 60.19% reported by the Office of the Baccalauréat of Cameroon (OBC) the same year [6], likely due in-part to the lower-performing nature of schools in our sample.

Our measured rate may also differ from that reported by the OBC due to the inclusion of students who did not take the exam in the denominator. On publicly-posted pass lists, students' names were not shown if they fail the bac, so our pass rate could not distinguish between students who failed the bac and students who simply did not take the exam at all. Although this dilutes any possible measured effect size, for the purposes of this analysis, we consider the two non-passing cases to be the same. This is because, in either case, the student did not obtain the diploma that would be valuable for future career and educational prospects.

Of the 657 total students who passed in our sample, 172 students (34.40%) came from the practice test treatment. 167 students (33.40%) passed in the positive messaging group. 318 students (34.42%) passed in the control group. Due to the rank nature of the dependent variable, we use a Kruskal-Wallis, one-way, non-parametric test [41] to compare these groups. This test did not show any statistically significant differences between the groups (h = 0.1679, p = 0.9195). This suggests that students in all groups experienced the same pass rate.

6 DISCUSSION

We discuss the implications of our results in relation to past work on practice tests and work using SMS in developing contexts. In particular, we describe how students' experiences and usage of our practice tests did or did not match our design intentions. We consider the challenges we faced that relate to our use of SMS for studying. Finally, we describe avenues for future research.

6.1 Extending Understanding of Practice Tests

In creating our quiz intervention, we aimed to embed specific benefits of practice testing into its design. Students' experiences both matched these intended benefits and included unexpected advantages. We discuss how our results provide more detail on the mechanisms behind the learning benefits of practice testing.

6.1.1 Testing Effect by Strengthening Knowledge Structures. We found that practice tests can prompt students to recall prior learning experiences even if the quizzes did not match students' own curriculum. This is important because most past research on the testing effect has assumed that the party providing practice questions is also the one determining the content on the summative exam. Past work made use of practice and summative tests written by the researcher [15, 50, 64], an individual teacher [19, 23, 38, 46], or generated by the researcher from a teacher's materials [48].

For some of the most important exams in a student's career, this assumption is not valid. Standardized tests, such as the baccalauréat, are written or curated by a centralized testing board, and teachers attempting to prepare their students for this test can only make educated guesses at the exact contents of the exam. We find that practice tests can still benefit students in these contexts by cueing students to review related material, potentially supporting the claim that practice tests can activate associated memories to help students answer questions that were never directly learned [12, 17]. Challenging questions are also more likely to prompt additional study, which might explain why more difficult practice tests can lead to stronger test effects than easy tests [40].

To successfully review related material requires the student to have a strong understanding of what constitutes "related material." Students must have some existing familiarity with the domain of the target material. Beyond simple memory operations, such as information retrieval, practice tests also induce students to evaluate how question content relates to their existing knowledge of the domain. By doing so, practice tests can enhance related retrieval structures to improve the learner's skill and mastery over that domain of knowledge [14, 18]. This effect may be related to improving students' organization of concepts, a benefit of practice tests observed by previous researchers [79].

6.1.2 Formative Self-Assessment and Study Strategies. We find that practice tests can serve to provide students with formative assessment, in agreement with past literature [23, 46, 51]. Students used our quizzes to identify shortcomings and knowledge gaps that they described to be helpful to focusing their preparations for the baccalauréat. We also noticed that students had different strategies for allocating their time to different study material, and these strategies could be influenced by formative assessment. This suggests that the benefits of formative assessment are uneven, depending on the review strategies of the particular student.

6.1.3 Creating Distributed Classroom Environments. We intended our practice tests to prompt study on an evenly spaced schedule to help students retain learned material [38, 46, 47]. Our results supported this, as students used the quizzes to focus their studies at the times of their receipt, but that studying was not necessarily done as individuals. Instead, our SMS-based quizzes served a social role in connecting students with each other. The automated system produced a synchronized schedule of questions, where classmates would be working on the same question at the same time. Some students shared messages with their friends only to realize later that their friends had already received the same questions.

Several students in our focus groups described sharing quiz questions, working on them collaboratively, and comparing notes on previously answered questions, whether through physical study groups or in web-based chat rooms. While the quiz intervention did not necessarily create these study groups or chat rooms, the shared materials and synchronicity of the SMS-based intervention generated the environment of a distributed classroom, where students had similar studying experiences and a shared social grounding despite not regularly attending classes together during the month of independent study in May.

Past work on practice tests outside of the classroom have avoided sharing and focused on academic integrity to ensure students engage seriously with the practice material [20]. However, in addition to added benefits as a social function, sharing the practice material could potentially be a sign of a deeper level of engagement than keeping the material to one's self. This is because students are less likely to share material that is less interesting and because shared material can lead to students reviewing related material with their collaborators. Students did not see sharing quiz material as "cheating" but the best way to make it into a learning experience.

6.2 Future of SMS in Mixed-Media

The experience of using SMS for practice testing was a mixed bag. Our intervention had benefits related to its underlying medium, namely the ability to maintain a regular schedule to prompt students to study and create a distributed classroom environment. These affordances of SMS-based practice tests cannot be entirely replicated in paper or web-based mediums. However, students also felt limited by SMS's text-only nature and were frustrated by the lack of more detailed feedback that would have been easier with other mediums. Some students explicitly suggested that we instead build a more multimedia-oriented digital platform.

However, moving away from SMS is not necessarily without costs due to how prevalent basic phones are in developing contexts. Past work has found that even mobile instant messaging (MIM)

platforms, such as WhatsApp, are not drop-in replacements for SMS due to their need for smartphones [60]. In contexts where smartphones are not widespread or well-accepted as educational tools, we suggest that a mixed media approach may be most beneficial. For example, mixing paper-based practice test materials with SMS-based scheduling and reminders may allow students to capture the unique benefits of both modes.

6.3 Limitations

Our work included several limitations related to the measurement of baccalauréat pass rates. One of the biggest limitations was the nature of bac pass lists which required considerable effort to gather from testing locations throughout Yaoundé. Digitization was also challenging due to the lack of a common identifier and the need to instead match results by student name. Though we focused our efforts for the highest coverage of students in our samples and used double-entry to reduce errors, there is still a possibility that we missed some students when gathering results. Finally, the binary pass-or-fail nature of pass lists and the lack of visibility on whether students failed or did not take the bac greatly reduced the ability of our statistical analysis to detect any effects.

In addition, as a field study conducted outside of a controlled environment, we could not enforce isolation between groups. Though we sent practice test content to students' own phones outside of school hours, we had no way to prevent students from sharing with their classmates, who may be in other study conditions. In fact, some of our participants explicitly described doing so. Further, though some students reported using the quiz passively, we had no way to measure how prevalent or deep this engagement was in comparison to active participation.

Despite these limitations, the bac is still an interesting instrument to measure student learning. It is a real-life exam, evaluated by an independent grading board, with great consequences to our students. Studying the testing effect on standardized exams is pertinent to real educational outcomes, but gathering results on actual standardized exams can be difficult due to the privacy sensitivity of such results. This was only possible for our work due to the public nature of pass lists.

6.4 Future Work

Our work suggests several directions for future research. One direction relates to practice tests and the testing effect for standardized exams such as the bac. Given the prevalence of standards-based assessment in many formal education systems, students in Cameroon are not the only ones whose educational futures are tied to a high-stakes standardized exam. Future work could examine the effect of practice testing on similar exams such as the SAT and other national exit or admission exams. As students' preparations for these exams occur over longer periods of times, effective use of practice testing may also need to occur over a similar duration. Analyzing exam outcomes at a more granular level than crude pass or fail would provide more visibility into the impact of practice testing.

We also suggest several directions for study related to the different modes that students engaged with our quizzes. This includes collaborative study of practice test material. For example, future work on practice tests outside of the classroom is needed to distinguish between helpful and non-helpful aspects of collaboration and design materials to selectively encourage the former. Collaborative practice tests can also be explored within a classroom environment and doing so may also shed light on what, if any, are the benefits. Because high-participators and high-performers are often overlapping sets, designing explicitly collaborative interventions may encourage participation from students who would otherwise be less inclined but whom stand to benefit the most.

Another future direction for studying practice tests include active versus passive participation and what differences exist in their impacts. This is not limited to education but relevant for any interactive intervention that is delivered via any remote, push-based mechanism, such as SMS. Better understanding the benefits and limitations of passive participation can potentially lead to better designed interventions to improve the experience of passive participators. Designing interventions solely for passive participation may also be more cost-effective, which is relevant to resource-constrained settings in ICTD.

Our work was situated in Cameroon, a sub-Saharan developing nation. We encourage more work in educational technology to be situated in non-Western contexts. Work based only in the Global North cannot be assumed to be evenly applicable to the entire globe, and comparative work can produce valuable insights. In our study, features of the context resulted in a set of practice testing benefits being highlighted that were less visible in prior literature. A further and deeper inspection of Cameroonian students' learning styles and the role of education could contribute to the growing body of research in comparative pedagogy and comparative behavioral science. Future academics could also take a critical stance to interrogate the role that colonial educational systems have in a postcolonial setting such as Cameroon.

Finally, technology differences in Cameroon led us to explore the relative benefits and disadvantages of SMS for practice testing. Work using SMS has a long history in developing contexts, but we encourage future work to consider using SMS in conjunction with other media in order to overcome some of the limitations of an SMS-only interaction.

7 CONCLUSION

This paper describes an SMS-based practice test intervention that sought to help students prepare for the bac, a high-stakes standardized exit exam in Cameroon. We deployed the intervention for nine weeks with 500 students from 23 secondary schools, collecting data via focus groups with students, participation behaviors, and bac pass rates. We found that students used the practice tests for formative assessment and as a prompt to study that could help them recall previously learned material and improve knowledge of related material. Students also used the quizzes in a collaborative fashion that could signify a higher degree of engagement. We describe how automated and push-based mediums like SMS can lead to reliable feedback, enable passive participation, and create a sense of a distributed classroom environment. However, students found the text-based, short-length, and high-cost nature of SMS

limiting. We suggest directions for future work in SMS-based interventions and for practice testing more broadly on standardized exams, collaborative studying, and in non-Western contexts.

8 ACKNOWLEDGEMENTS

REFERENCES

[1] Haitham Abaza and Michael Marschollek. 2017. SMS education for the promotion of diabetes self-management in low & middle income countries: a pilot randomized controlled trial in Egypt. *BMC Public Health* 17, 1 (dec 2017), 962. https://doi.org/10.1186/s12889-017-4973-5

[2] Afrobarometer. 2015. Cameroon. http://www.afrobarometer.org/countries/cameroon-0

[3] Pooja K. Agarwal, Laura D'Antonio, Henry L. Roediger, Kathleen B. McDermott, and Mark A. McDaniel. 2014. Classroom-based programs of retrieval practice reduce middle school and high school students' test anxiety. *Journal of Applied Research in Memory and Cognition* 3, 3 (sep 2014), 131–139. https://doi.org/10.1016/j.jarmac.2014.07.002

[4] Simon D. Angus and Judith Watson. 2009. Does regular online testing enhance student learning in the numerical sciences? Robust evidence from a large data set. *British Journal of Educational Technology* 40, 2 (mar 2009), 255–272. https://doi.org/10.1111/j.1467-8535.2008.00916.x

[5] Muhammad Asim, Fahad Shamshad, Muhammad Awais, and Ali Ahmed. 2017. Introducing Data mining for Predicting trends in School Education of Pakistan: Preliminary results and future directions. In *Proceedings of the Ninth International Conference on Information and Communication Technologies and Development - ICTD '17*, Vol. Part F1320. ACM Press, New York, New York, USA, 1–5. https://doi.org/10.1145/3136560.3136597

[6] Investir au Cameroun. 2019. Le taux de réussite au baccalauréat de l'enseignement général au Cameroun passe de 52,28% en 2018 à 60,18% en 2019. https://www.investiraucameroun.com/education/2307-13026-le-taux-de-reussite-au-baccalaureat-de-l-enseignement-general-au-cameroun-passe-de-52-28-en-2018-a-60-18-en-2019

[7] Aaron S Benjamin and Jonathan Tullis. 2010. What makes distributed practice effective? *Cognitive Psychology* 61, 3 (nov 2010), 228–247. https://doi.org/10.1016/j.cogpsych.2010.05.004

[8] Robert a. Bjork. 1975. Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In *Information processing and cognition: The Loyola Symposium.* Chapter 5, 123–144.

[9] Richard E. Boyatzis. 1998. *Transforming Qualitative Information: Thematic Analysis and Code Development.* Sage Publications, Inc.

[10] Thomas Brothen and Cathrine Wambach. 2001. Effective Student Use of Computerized Quizzes. *Teaching of Psychology* 28, 4 (oct 2001), 292–294. https://doi.org/10.1207/S15328023TOP2804_10

[11] T Buchanan. 2001. The efficacy of a World-Wide Web mediated formative assessment. *Journal of Computer Assisted Learning* 16, 3 (dec 2001), 193–200. https://doi.org/10.1046/j.1365-2729.2000.00132.x

[12] Andrew C. Butler. 2010. Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 36, 5 (2010), 1118–1133. https://doi.org/10.1037/a0019902

[13] Andrew C. Butler and Henry L. Roediger. 2008. Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition* 36, 3 (apr 2008), 604–616. https://doi.org/10.3758/MC.36.3.604

[14] Guillermo Campitelli. 2015. Memory behavior requires knowledge structures, not memory stores. *Frontiers in Psychology* 6, NOV (nov 2015), 1–4. https://doi.org/10.3389/fpsyg.2015.01696

[15] Shana K. Carpenter, Harold Pashler, and Nicholas J. Cepeda. 2009. Using tests to enhance 8th grade students' retention of U.S. history facts. *Applied Cognitive Psychology* 23, 6 (sep 2009), 760–771. https://doi.org/10.1002/acp.1507

[16] Sunandan Chakraborty, Tamali Bhattacharya, Plaban K. Bhowmick, Anupam Basu, and Sudeshna Sarkar. 2007. Shikshak: An Intelligent Tutoring System Authoring tool for rural education. In *2007 International Conference on Information*

and Communication Technologies and Development. IEEE, 1–10. https://doi.org/10.1109/ICTD.2007.4937410

[17] Jason C. K. Chan, Kathleen B. McDermott, and Henry L. Roediger. 2006. Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General* 135, 4 (2006), 553–571. https://doi.org/10.1037/0096-3445.135.4.553

[18] William G. Chase and K. Anders Ericsson. 1982. Skill and Working Memory. In *Psychology of Learning and Motivation - Advances in Research and Theory.* Vol. 16. 1–58. https://doi.org/10.1016/S0079-7421(08)60546-0

[19] Vincent A. Cicirello. 2009. On the role and effectiveness of pop quizzes in CS1. *ACM SIGCSE Bulletin* 41, 1 (mar 2009), 286. https://doi.org/10.1145/1539024.1508971

[20] David B Daniel and John Broida. 2004. Using Web-Based Quizzing to Improve Exam Performance: Lessons Learned. *Teaching of Psychology* 31, 3 (jul 2004), 207–208. https://doi.org/10.1207/s15328023top3103_6

[21] Catalina M Danis, Jason B Ellis, Wendy a Kellogg, Hajo Van Beijma, Bas Hoefman, Steven D Daniels, and Jan-willem Loggers. 2010. Mobile Phones for Health Education in the Developing World : SMS as a User Interface. *ACM Dev'10: 1st Annual Symposium on Computing for Development* (2010). https://doi.org/10.1145/1926180.1926197

[22] Tawanna Dillahunt, Zengguang Wang, and Stephanie D. Teasley. 2014. Democratizing higher education: Exploring MOOC use among those who cannot afford a formal education. *International Review of Research in Open and Distance Learning* 15, 5 (2014), 177–196. https://doi.org/10.19173/irrodl.v15i5.1841

[23] John L. Dobson. 2008. The use of formative online quizzes to enhance class preparation and scores on summative exams. *Advances in Physiology Education* 32, 4 (dec 2008), 297–302. https://doi.org/10.1152/advan.90162.2008

[24] Philippe C. Duchastel and Ronald J. Nungester. 1982. Testing Effects Measured with Alternate Test Forms. *The Journal of Educational Research* 75, 5 (may 1982), 309–313. https://doi.org/10.1080/00220671.1982.10885400

[25] John Dunlosky, Katherine A. Rawson, Elizabeth J. Marsh, Mitchell J. Nathan, and Daniel T. Willingham. 2013. Improving Students' Learning With Effective Learning Techniques. *Psychological Science in the Public Interest* 14, 1 (jan 2013), 4–58. https://doi.org/10.1177/1529100612453266

[26] Lisa K. Fazio, Pooja K. Agarwal, Elizabeth J. Marsh, and Henry L. Roediger. 2010. Memorial consequences of multiple-choice testing on immediate and delayed tests. *Memory & Cognition* 38, 4 (jun 2010), 407–418. https://doi.org/10.3758/MC.38.4.407 arXiv:NIHMS150003

[27] Vanessa Frias-Martinez, Jesus Virseda, and Aldo Gomero. 2012. Mobilizing education: Evaluation of a Mobile Learning Tool in a Low-Income School. *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services - MobileHCI '12* (2012), 441. https://doi.org/10.1145/2371574.2371642

[28] International Monetary Fund. 2018. *World Economic Outlook: Growth Slowdown, Precarious Recovery.* Technical Report April. International Monetary Fund, Washington, DC.

[29] Kiran Gaikwad, Gaurav Paruthi, and William Thies. 2010. Interactive DVDs as a platform for education. In *Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development - ICTD '10.* ACM Press, New York, New York, USA, 1–10. https://doi.org/10.1145/2369220.2369233

[30] Arthur I. Gates. 1917. Recitation as a factor in memorizing. *Archives of Psychology* 40 (1917), 104.

[31] Christopher J. Gill, Bao Le Ngoc, Nafisa Halim, Ha Nguyen Viet, Anna Larson Williams, Nguyen Van Tan, Marion McNabb, Lien Tran Thi Ngoc, Ariel Falconer, Hai An Phan Ha, Julia Rohr, Hai Hoang, James Michiel, Tam Nguyen Thi Thanh, Liat Bird, Hoang Pham Vu, Mahlet Yeshitla, Ha Van Nhu, and Lora Sabin. 2016. The mCME Project: A Randomized controlled trial of an SMS-Based continuing medical education intervention for improving medical knowledge among Vietnamese community based physicians' assistants. *PLoS ONE* 11, 11 (2016), 1–19. https://doi.org/10.1371/journal.pone.0166293

[32] Robert B. Graham. 1999. Unannounced Quizzes Raise Test Scores Selectively for Mid-Range Students. *Teaching of Psychology* 26, 4 (oct 1999), 271–273. https://doi.org/10.1207/S15328023TOP260406

[33] Joseph Henrich, Steven J. Heine, and Ara Norenzayan. 2010. The weirdest people in the world? *Behavioral and Brain Sciences* 33, 2-3 (jun 2010), 61–83. https://doi.org/10.1017/S0140525X0999152X

[34] Nick Hughes and Susie Lonie. 2007. M-PESA: Mobile Money for the "Unbanked" Turning Cellphones into 24-Hour Tellers in Kenya. *Innovations: Technology, Governance, Globalization* 2, 1-2 (2007), 63–81. https://doi.org/10.1162/itgg.2007.2.1-2.63

[35] Stephen S. Intille. 2004. Ubiquitous computing technology for just-in-time motivation of behavior change. *Studies in health technology and informatics* 107 (2004), 1434–7. https://doi.org/10.3233/978-1-60750-949-3-1434

[36] Simy Joy and David A. Kolb. 2009. Are there cultural differences in learning style? *International Journal of Intercultural Relations* 33, 1 (jan 2009), 69–85. https://doi.org/10.1016/j.ijintrel.2008.11.002

[37] Peter Kelly. 2013. Comparative Pedagogy: Making Sense of Cultural Complexity. *Research in Comparative and International Education* 8, 4 (dec 2013), 415–427. https://doi.org/10.2304/rcie.2013.8.4.415

[38] Frank M. Kika, T. F. McLaughlin, and J. Dixon. 1992. Effects of Frequent Testing of Secondary Algebra Students. *The Journal of Educational Research* 85, 3 (jan 1992), 159–162. https://doi.org/10.1080/00220671.1992.9944432

[39] Grace Kong, Daniel M. Ells, Deepa R. Camenga, and Suchitra Krishnan-Sarin. 2014. Text messaging-based smoking cessation intervention: A narrative review. *Addictive Behaviors* 39, 5 (2014), 907–917. https://doi.org/10.1016/j.addbeh.2013.11.024

[40] Nate Kornell, Matthew Jensen Hays, and Robert A. Bjork. 2009. Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 35, 4 (2009), 989–998. https://doi.org/10.1037/a0015729

[41] William H Kruskal and W Allen Wallis. 1952. Use of Ranks in One-Criterion Variance Analysis. *J. Amer. Statist. Assoc.* 47, 260 (dec 1952), 583. https://doi.org/10.2307/2280779

[42] Neha Kumar, Pallavi Lal, Richard Anderson, Trevor Perrier, Michelle Desmond, Kiersten Israel-Ballard, Vikrant Kumar, Sudip Mahapatra, Anil Mishra, Shreya Agarwal, and Rikin Gandhi. 2015. Projecting Health: Community-Led Video Education for Maternal Health. In *Proceedings of the Seventh International Conference on Information and Communication Technologies and Development - ICTD '15*. ACM Press, New York, New York, USA, 1–10. https://doi.org/10.1145/2737856.2738023

[43] Shaimaa Lazem and Hussein Aly Jad. 2017. We play we learn: Exploring the value of digital educational games in Rural Egypt. *Conference on Human Factors in Computing Systems - Proceedings* 2017-May (2017), 2782–2791. https://doi.org/10.1145/3025453.3025593

[44] Kathleen Luchini, Chris Quintana, and Elliot Soloway. 2004. Design guidelines for learner-centered handheld tools. In *Proceedings of the 2004 conference on Human factors in computing systems - CHI '04*, Vol. 6. ACM Press, New York, New York, USA, 135–142. https://doi.org/10.1145/985692.985710

[45] Carleen Maitland and Eric Obeysekare. 2015. The creation of capital through an ICT-based learning program. *Proceedings of the Seventh International Conference on Information and Communication Technologies and Development - ICTD '15* (2015), 1–10. https://doi.org/10.1145/2737856.2738024

[46] Michael Marcell. 2008. Effectiveness of Regular Online Quizzing in Increasing Class Participation and Preparation. *International Journal for the Scholarship of Teaching and Learning* 2, 1 (jan 2008). https://doi.org/10.20429/ijsotl.2008.020107

[47] V. T. Mawhinney, D. E. Bostow, D. R. Laws, G. J. Blumenfeld, and B. L. Hopkins. 1971. A comparison of students studying-behavior produced by daily, weekly, and three-week testing schedules. *Journal of Applied Behavior Analysis* 4, 4 (1971), 1310703. https://doi.org/10.1901/jaba.1971.4-257

[48] Mark A. McDaniel, Pooja K. Agarwal, Barbie J. Huelser, Kathleen B. McDermott, and Henry L. Roediger. 2011. Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology* 103, 2 (2011), 399–414. https://doi.org/10.1037/a0021782

[49] Mark A. McDaniel, Janis L. Anderson, Mary H. Derbish, and Nova Morrisette. 2007. Testing the testing effect in the classroom. *European Journal of Cognitive Psychology* 19, 4-5 (jul 2007), 494–513. https://doi.org/10.1080/09541440701326154

[50] Mark A. McDaniel and Ronald P. Fisher. 1991. Tests and test feedback as learning sources. *Contemporary Educational Psychology* 16, 2 (apr 1991), 192–201. https://doi.org/10.1016/0361-476X(91)90037-L

[51] Mark A. McDaniel, Kathleen M. Wildman, and Janis L. Anderson. 2012. Using quizzes to enhance summative-assessment performance in a web-based class: An experimental study. *Journal of Applied Research in Memory and Cognition* 1, 1 (mar 2012), 18–26. https://doi.org/10.1016/j.jarmac.2011.10.001

[52] Kathleen B. McDermott, Pooja K. Agarwal, Laura D'Antonio, Henry L.I. Roediger, and Mark A. McDaniel. 2014. Both multiple-choice and short-answer quizzes enhance later exam performance in middle and high school classes. *Journal of Experimental Psychology: Applied* 20, 1 (2014), 3–21. https://doi.org/10.1037/xap0000004

[53] Laura McSweeney and Joan Wyzkoski Weiss. 2003. Assessing the math online tool: A progress report. *Mathematics and Computer Education* 37, 3 (2003), 348–357.

[54] Maletsabisa Molapo and Gary Marsden. 2013. Software support for creating digital health training materials in the field. *Proceedings of the Sixth International Conference on Information and Communication Technologies and Development Full Papers - ICTD '13 - volume 1* (2013), 205–214. https://doi.org/10.1145/2516604.2516632

[55] Fidele Ngabo, Judith Nguimfack, Friday Nwaigwe, Catherine Mugeni, Denis Muhoza, David R. Wilson, John Kalach, Richard Gakuba, Corrine Karema, and Agnes Binagwaho. 2012. Designing and Implementing an Innovative SMS-based alert system (RapidSMS-MCH) to monitor pregnancy and reduce maternal and child deaths in Rwanda. *The Pan African medical journal* 13 (2012), 31. http://www.ncbi.nlm.nih.gov/pubmed/23330022http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3542808

[56] Richard E. Nisbett. 2003. *The geography of thought: How Asians and Westerners think differently ... and why.* Free Press.

[57] Harold Pashler, Nicholas J Cepeda, John T Wixted, and Doug Rohrer. 2005. When Does Feedback Facilitate Learning of Words? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31, 1 (2005), 3–8. https://doi.org/10.1037/0278-7393.31.1.3

[58] Udai Pawar, Joyojeet Pal, and Kentaro Toyama. 2006. Multiple Mice for Computers in Education in Developing Countries. In *2006 International Conference on Information and Communication Technologies and Development*. IEEE, 64–71. https://doi.org/10.1109/ICTD.2006.301840

[59] Trevor Perrier, Nicola Dell, Brian DeRenzi, Richard Anderson, John Kinuthia, Jennifer Unger, and Grace John-Stewart. 2015. Engaging Pregnant Women in Kenya with a Hybrid Computer-Human SMS Communication System. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15* (2015), 1429–1438. https://doi.org/10.1145/2702123.2702124

[60] Anthony Poon, Sarah Giroux, Parfait Eloundou-Enyegue, François Guimbretiere, and Nicola Dell. 2019. Engaging High School Students in Cameroon with Exam Practice Quizzes via SMS and WhatsApp. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. ACM Press, New York, New York, USA, 1–13. https://doi.org/10.1145/3290605.3300712

[61] Fraser J.M. Reid and Donna J. Reid. 2010. The expressive and conversational affordances of mobile messaging. *Behaviour & Information Technology* 29, 1 (jan 2010), 3–22. https://doi.org/10.1080/01449290701497079

[62] Ruth Rettie. 2009. SMS: Exploiting the Interactional Characteristics of Near-Synchrony. *Information, Communication & Society* 12, 8 (dec 2009), 1131–1148. https://doi.org/10.1080/13691180902786943

[63] Henry L. Roediger and Andrew C. Butler. 2011. The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences* 15, 1 (jan 2011), 20–27. https://doi.org/10.1016/j.tics.2010.09.003

[64] Henry L. Roediger and Jeffrey D. Karpicke. 2006. Test-Enhanced Learning. *Psychological Science* 17, 3 (mar 2006), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

[65] Henry L. Roediger and Elizabeth J. Marsh. 2005. The Positive and Negative Consequences of Multiple-Choice Testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31, 5 (2005), 1155–1159. https://doi.org/10.1037/0278-7393.31.5.1155

[66] Henry L. Roediger III, Adam L. Putnam, and Megan A. Smith. 2011. Ten Benefits of Testing and Their Applications to Educational Practice. In *Psychology of Learning and Motivation - Advances in Research and Theory*. Vol. 55. Elsevier Inc., 1–36. https://doi.org/10.1016/B978-0-12-387691-1.00001-6

[67] Amy M Shapiro and Leamarie T Gordon. 2013. Classroom clickers offer more than repetition: Converging evidence for the testing effect and confirmatory feedback in clicker-assisted learning. *Journal of Teaching and Learning with Technology* 2, 1 (2013), 15–30.

[68] Melissa L. Shirley and Karen E. Irving. 2015. Connected Classroom Technology Facilitates Multiple Components of Formative Assessment Practice. *Journal of Science Education and Technology* 24, 1 (feb 2015), 56–68. https://doi.org/10.1007/s10956-014-9520-x

[69] Elliot Soloway, Nancy Scala, Shari L. Jackson, Jonathan Klein, Chris Quintana, James Reed, Jeff Spitulnik, Steven J. Stratford, Scott Studer, and Jim Eng. 1996. Learning theory in practice: case studies of learner-centered design. In *Proceedings of the SIGCHI conference on Human factors in computing systems common ground - CHI '96*. ACM Press, New York, New York, USA, 189–196. https://doi.org/10.1145/238386.238476

[70] Lisa K. Son and Nate Kornell. 2008. Research on the Allocation of Study Time. In *Handbook of Metamemory and Memory*. Routledge, 333–351. https://doi.org/10.4324/9780203805503.ch17

[71] Caroline Stratton, Dan Sholler, Diane Bailey, Paul Leonardi, and Carlos Rodríguez-Lluesma. 2016. Competing institutional logics in ICT4D education projects: A South American study. In *Proceedings of the Eighth International Conference on Information and Communication Technologies and Development - ICTD '16*, Vol. 03-06-June. ACM Press, New York, New York, USA, 1–11. https://doi.org/10.1145/2909609.2909665

[72] Mark G. Urtel, Rafael E. Bahamonde, Alan E. Mikesky, Eileen M. Udry, and Jeff S. Vessely. 2006. On-line Quizzing and it's Effect on Student Engagement and Academic Performance. *The Journal of Scholarship of Teaching and Learning* 6, 2 (2006), 84–92.

[73] Rama Adithya Varanasi, René F Kizilcec, and Nicola Dell. 2019. How Teachers in India Reconfigure their Work Practices around a Teacher-Oriented Technology Intervention. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (nov 2019), 1–21. https://doi.org/10.1145/3359322

[74] Michael E. W. Varnum, Igor Grossmann, Shinobu Kitayama, and Richard E. Nisbett. 2010. The Origin of Cultural Differences in Cognition: Evidence for the Social Orientation Hypothesis. *Current Directions in Psychological Science* 19, 1 (feb 2010), 9–13. https://doi.org/10.1177/0963721409359301

[75] Jan Visser and John M. Keller. 1990. The clinical use of motivational messages: an inquiry into the validity of the ARCS model of motivational design. *Instructional Science* 19, 6 (1990), 467–500. https://doi.org/10.1007/BF00119391

[76] Richard Joseph Waddington and SungJin Nam. 2014. Practice Exams Make Perfect: Incorporating Course Resource Use into an Early Warning System. In *Proceedins*

of the Fourth International Conference on Learning Analytics And Knowledge - LAK '14. ACM Press, New York, New York, USA, 188–192. https://doi.org/10.1145/2567574.2567623

[77] Ebenezer L. Wirba, Boniface N. Epo, and Francis M. Baye. 2013. Explaining Growth in Household Real Food Consumption Expenditure in Cameroon: 2001-2007. *Tanzania Economic Review* 3, 1-2 (2013), 17–34.

[78] Susan Wyche, Nightingale Simiyu, and Martha E. Othieno. 2019. Understanding women's mobile phone use in rural Kenya: An affordance-based approach. *Mobile Media & Communication* 7, 1 (jan 2019), 94–110. https://doi.org/10.1177/2050157918776684

[79] Franklin M. Zaromb and Henry L. Roediger. 2010. The testing effect in free recall is associated with enhanced organizational processes. *Memory & Cognition* 38, 8 (dec 2010), 995–1008. https://doi.org/10.3758/MC.38.8.995