Developing a prototype web-app for numeracy assessment and teaching

UROP Internship Report, September 2018

Zhilin Wang, Andrew Caines, Ianthi Tsimpli
University of Cambridge

Abstract

We propose a web-based application ('web-app') for primary-aged students in India to develop verbal reasoning, problem solving and English literacy. Solving word problems in mathematics depends on not only students' understanding of mathematical concepts but also their ability to linguistically understand what the question is trying to ask. In an example question, Jake runs at 2 metres per seconds. If Jane is _____ than Jake by 1 metre per second, at what speed is Jane moving at? It is not possible for us to answer this question if we do not understand the meaning of the word "faster", which was blanked out above. This situation, however, is not uncommon when word problems are posed in a language that is not of the child's native language. This is the struggle faced by many Indian children who learn in English but speak other languages as their mother tongue.

To improve their learning of mathematics and make more accurate assessment of their grasp of numerical concepts, we designed a web-app that seeks to provide scaffolding for child learners. The web-app has three main features that are pedagogically significant. The first is the ability for the user to view dictionary definitions for complex English words. The second is for users who require more guidance to translate the entire question into their native language. The last is for the user to be able to request hints and checks on their responses before submission. Their use of these features is also logged for researchers to analyze, which can in turn help to personalize such guidance in the future. Our web-app is a scalable solution that currently includes a thousand questions from an open-source dataset and can be customized to include questions from an external source. Future work includes personalizing the experience based on the mathematical attainment of the student and converting the web-app into an offline Android app that is useful in regions where Internet connections are hard to find and/or unstable.

Introduction

We proposed the development of a web-based application (a 'web-app') for the assessment teaching of verbal reasoning, problem-solving and English literacy to primary school children in India aged 9-10 years. Verbal reasoning questions are aimed at testing the user's numeracy skills through verbal puzzles – often referred to as 'meta-maths' and 'word problems'.

However, since the questions are by definition linguistically expressed, a child's literacy level produces a confound, potentially obfuscating their true numeracy level if their literacy is not good enough to understand the question.

We chose India as the target of this web-app because (a) we have experience working in the region, and (b) demand for literacy in English is especially high in the Indian school system, as evidenced by the increase in (low-cost) English-medium instruction schools across the country. Although many schools teach in English, teachers' English proficiency is not universally high and learning outcomes can be lower than expected given the age of the children and their years of schooling. We therefore intend to provide a web-app with linguistic scaffolding: both dictionary definitions for complex English words, and text translations into Indian languages such as Hindi and Telugu with the expectation that the facility to read a question in the child's mother tongue(s) will help them understand the question and demonstrate their numeracy skills.

Examples of a word-problem (1) and meta-maths question (2) are presented below:

(1) Kerosene comes in 5 litre cans. Ashoka needs 17 litres of kerosene for the household. How many cans must he buy: 5, 2, 8 or 4?

(2) Here is a page from Seema's maths notebook.
5 x 4 = 9
3 x 2 = 5
4 x 2 = 6
Why does Seema make these mistakes?

The web-app presents tests and teaching materials to learners in an interactive fashion, and will eventually allow for video lessons to accompany the questions, so that the learner can independently learn to solve the questions they find difficult.

During this UROP internship, the first author developed a prototype web-app for numeracy teaching and testing. In this report he explains progress so far and potential directions for future work.

Web-app development

The development process had two main stages: the first was to build a minimal prototype to test out the feasibility of each pedagogic feature we had in mind while the second was to scale up our prototype by incorporating open-sourced question sets into the web-app.

First Stage: Prototype development

To create the web-app, we used the Python Flask micro-framework to set up a web-server. We then hosted it on pythonanywhere.com, which is viewable at zw322.pythonanywhere.com. Python Flask was chosen because it was easy to prototype with and has much room for customization, which is necessary for this use-case.

Each question consists of a question text, sometimes multiple options for users to choose from as well as a supporting material in the form of a table or an image. This is stored in a SQLite database and an example of it is shown in Figure 1.

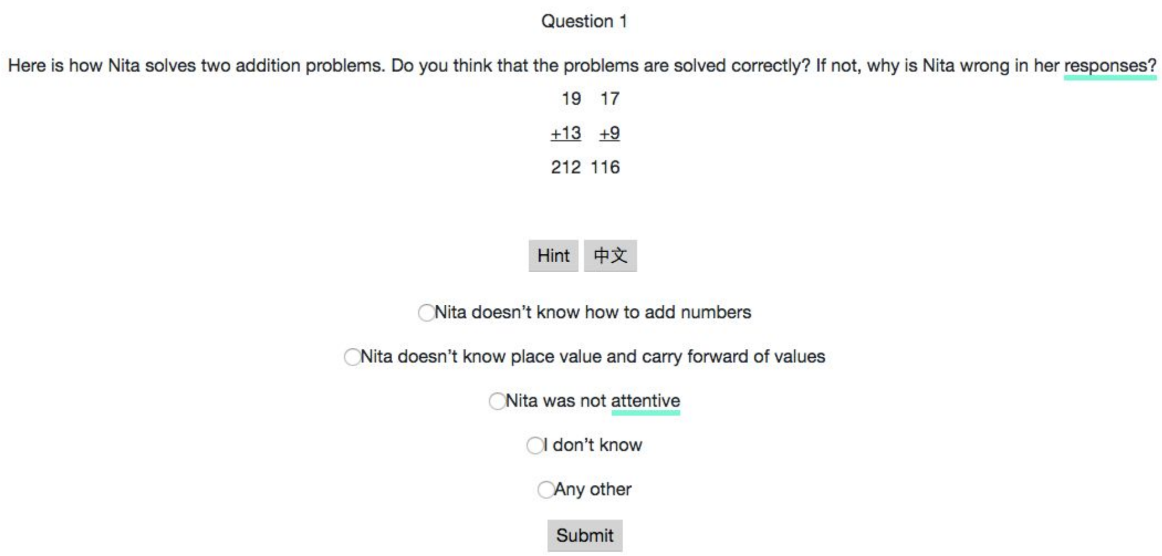

Figure 1. Question with supporting table and options

The first pedagogic intervention we implemented was to provide dictionary definitions of certain difficult words in the question. This is important because students might find questions challenging not due to poor understanding of mathematical concepts but the incompleteness of question comprehension. In figure 2 below, should a student not understand the word “represent”, he/she will not able to satisfy the requirements of the question even if he/she appreciates the underlying mathematical concept behind the question. To alleviate the issue, we designed a system that reveals meanings of words when students click on the words they find hard to understand. Because there is no objective metric for how difficult a word is to a English as a foreign language learner, we relied upon online datasets that asked users to identify and rate complex words in the sentence [Yimam et al 2018 https://aclanthology.info/papers/W18-0507/w18-0507]. The dataset contains sentences from WikiNews, Wikipedia and general News articles. Because our target users are mostly aged 9 – 10, we decided to have a very low boundary on what we consider to be a difficult word – including every word that has been labeled as difficult by at least one rater.

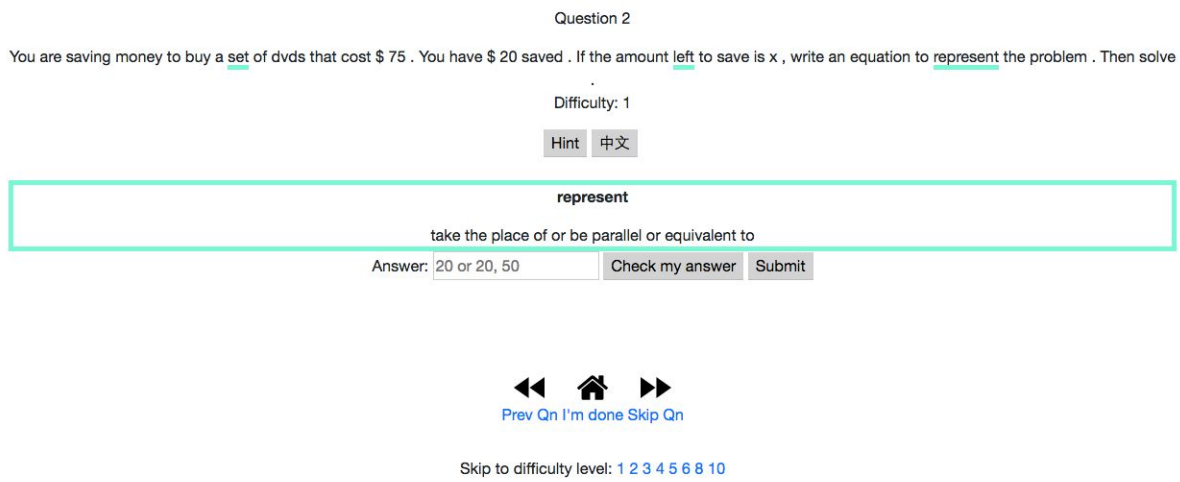

Figure 2. Use of definitions of difficult words

The second pedagogic intervention we built was to allow students to use hints and to check their answers before submitting. We saw the possibility of such hints being encouraging to the student when he/she is faced with a challenging question. The hint can remind the student of the specific concept he/she needs in order to answer the question. This hint can come in a multitude of formats such as a partial solution, a more detailed breakdown of the question or as in Figure 3 below, a video explaining the concept needed. As a demonstration, we chose to use YouTube videos because there is a lot of diversity in content, with many explaining topics in engaging manners. Allowing students to check their answers against the true answers gives students an opportunity to re-examine their working in case they obtained a mistake in their response due to careless oversights.

Figure 3. Providing students hints when required

The third and final pedagogic intervention we designed is a translation tool. We thought that the wholesale translation of the question was necessary to meet the needs of students with minimal proficiency in the English language. Even for them, there is likely to be an improvement in English literacy because exposing them to the English version of the question prior to translating might help them to acquire understanding of basic vocabulary over time. When the student clicks on a button, the entire question will be translated into the student’s mother tongue as shown in Figure 4. This was done by calling Google translate API and storing the translation in a file. Figure 4 shows the translation of the question in Figure 2 into Chinese, which the first author was familiar with and could hence verify the accuracy of the translation. For practical use, Telugu and Hindi translations have been included in the application and be easily put into action by minor adjustment in the codebase (See README.md in the GitHub repository https://github.com/Zhilin123/urop for details).

Question 2

你正在省钱买一套售价75美元的dvds。你节省了20美元。如果要保存的金额是x，请写一个等式来表示问题。然后解决。

Difficulty: 1

Figure 4: Translating question into Chinese

Second Stage: Scale up

After demonstrating the basic features of the web-application, our next step was to increase the number of questions available for the student to work on. We explored different ideas about how we can obtain the questions but finally decided that using an open source dataset of math problems was the most suitable. This decision was made on the basis that the questions ultimately used could be adjusted to the needs of the students/researchers. This means that our aim in the web-app development phase is to demonstrate how our application can be used with a large problem set. Therefore rather than spending much effort creating a proprietary dataset, we can simply use the ones that are publicly available.

The dataset we decided to use was a set of 1000 questions prepared by Kushman et. al. (2014) [Kushman et al 2014 http://aclweb.org/anthology/P14-1026]. Briefly, this dataset was prepared by crowdsourcing word problems from algebra.com, which was then manually labeled with solution equation(s) and numeric answer(s). We chose this dataset over others such as Microsoft’s Dolphin 18K dataset because this dataset contains word problems concerning mathematics used in personal contexts rather than mathematics out of context. For example, compare the question in Figure 2 to this question from the Dolphin 18K dataset "The sum of two numbers is 18 and their difference is 4. What are the two numbers?"

We built upon the dataset to categorize questions into 5 different categories based on the type of mathematical concepts required to answer them. For students, this might be helpful for them to practise and sharpen their grasp of specific mathematical concepts.

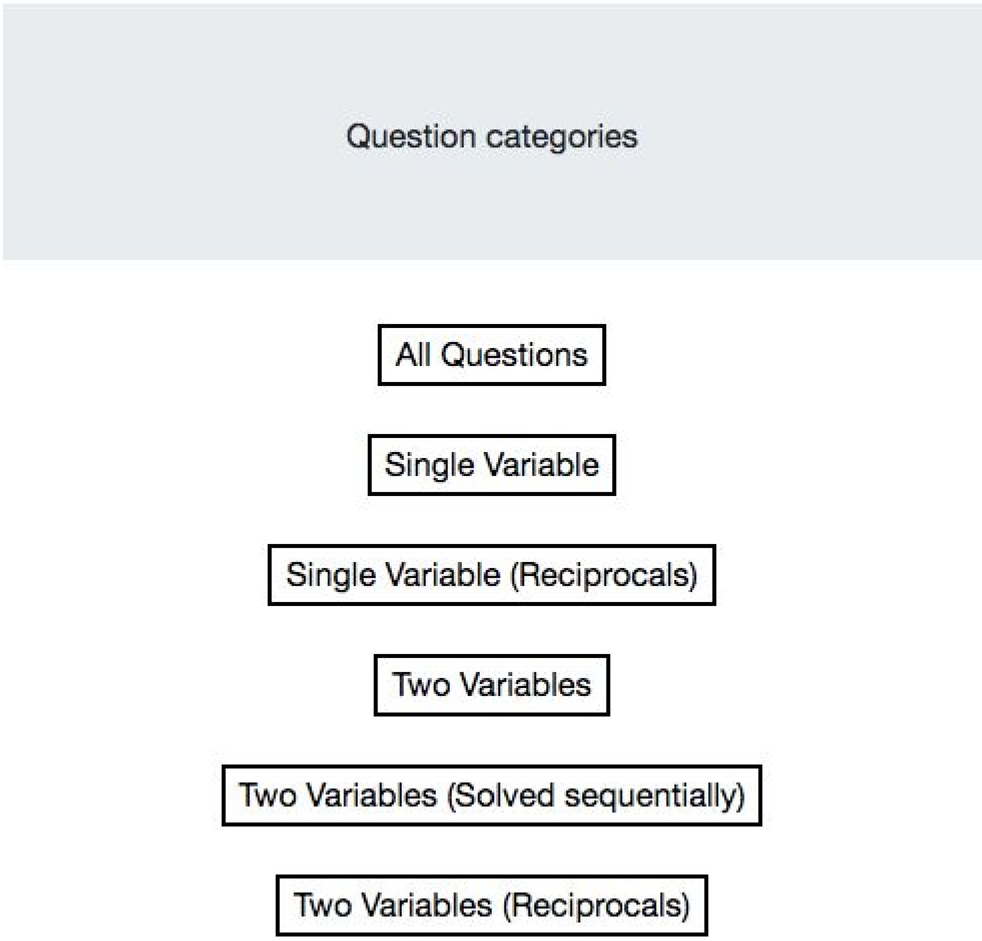

Figure 5. Question Categories on Front Page

As students are working through questions, useful information is also collected to facilitate subsequent research. The information collected include the question attempted, the student's answer(s) as well as whether they used the options for retrieving dictionary definitions, translations and hints. We also log the session ID in order to analyse how students' performance changes over different sittings. Not only will the information collected help in research, the user's behaviour in utilizing the various scaffolding features we provide can help us to improve and personalize user experiences in the future.

Finally, the significant number of questions in our database made it useful to implement a few additional aspects to improve the user experience. The first is to be able to continue on from the last question the student attempted. In this way, the student will not have to restart from the beginning each time they finish a session. The second is the ability to know how many questions they got right in the previous session as well as in all time. This is helpful in informing students whether they have understood a concept comprehensively or should spend more time understanding the underlying concepts before investing time in further practice. The third, as shown in Figure 6, is to be able to compare their own performance to that of their friends. We felt that this can provide some form of gamified extrinsic motivation that can encourage students to practise more frequently. Knowing which of their friends performs better can also inform them who to turn to when they need to seek guidance from someone else.

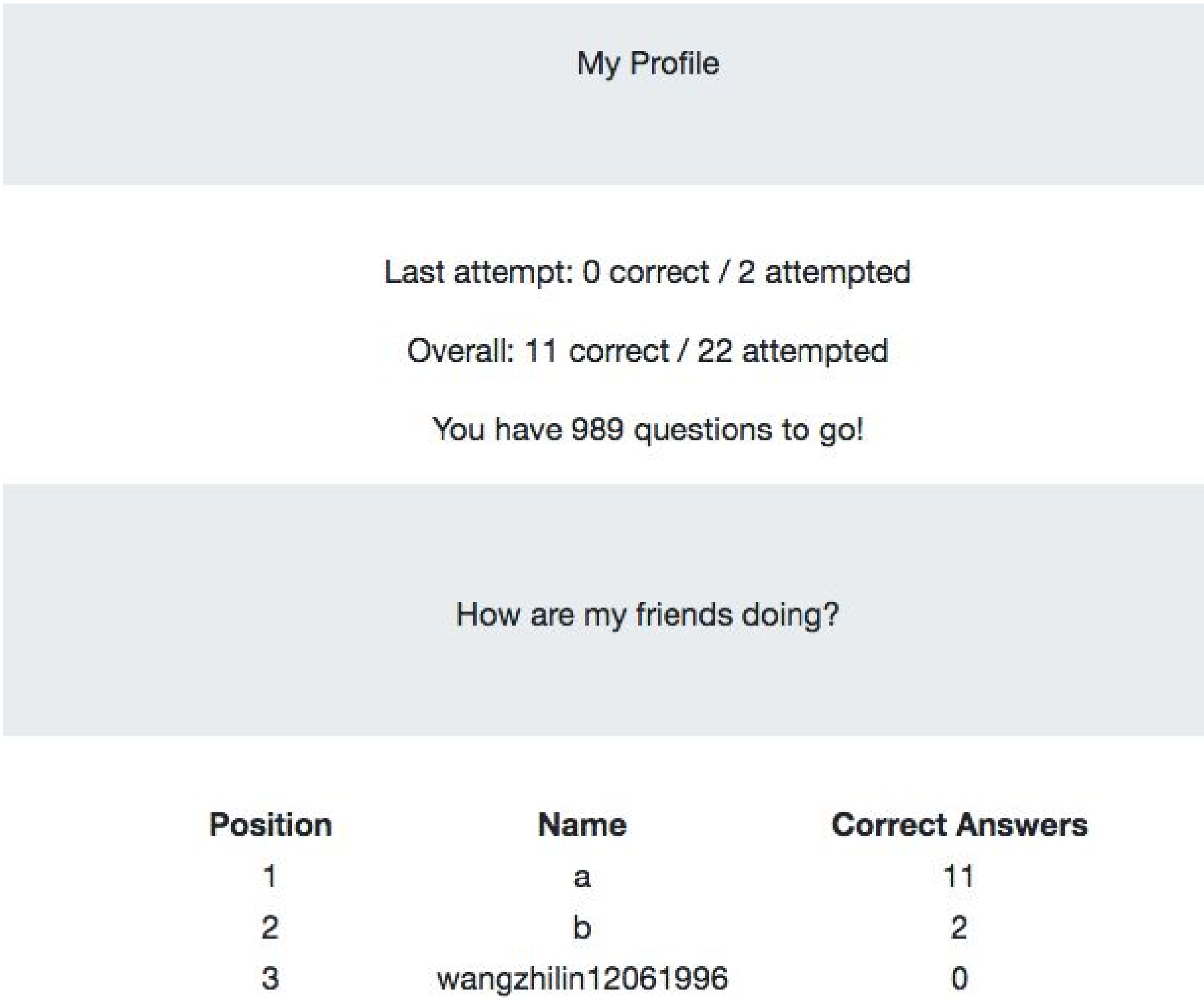

Figure 6. User profile and Leaderboard Among Friends

Conclusions & Future work

In this project we aim to develop a web-app for mathematics testing and teaching. We specifically focus on word problems and meta-maths questions which challenge students' literacy as well as numeracy levels. In order to support students whose literacy level may hinder their ability to demonstrate their true numeracy level, we introduce scaffolding features in the app -- dictionary definitions for complex words, full text translation into Chinese, Hindi and Telugu, and teaching videos to provide more background to the given numeracy problem. It is our hypothesis that such features will help in certain scenarios such as those found in India, where schooling is carried out in English -- a second language for many Indian children. We can test this hypothesis through randomised controlled trials involving pre-tests, an intervention with the web-app (or control), and post-tests.

However, the app first requires more content, user testing and possible refinement before deployment in research fieldwork. In addition we will need to verify the data we wish to collect, as this may require adjustments to the app and logging procedure. Thirdly we must apply for further funding and ethics approval for this study.

One avenue for future work is to introduce a personalisation element to the web-app. For this to be effective we need to add more content and build what amounts to a curriculum covering at least one year's schooling for the target age group. Possible methods to do so include manual curation, ingesting content already available on the Internet, or collaborating with a publisher of mathematics textbooks (such as Cambridge University Press) to obtain and include their content in the app.

In the personalisation scenario, on first use of the web-app students will undertake a placement test in order to be assessed and assigned to a proficiency level. The web-app would then guide the student through the course curriculum whilst undertaking continuous assessment to verify they are receiving appropriate guidance. It would thus become an example of an adaptive, automated teaching system, in which students' pathways through the curriculum are individualised and determined by their assessed proficiency at any given point.

Another potential direction is to convert the web-app into an Android mobile app that can be used offline. This might be useful because some areas where research might be carried out do not have stable wifi or mobile networks. For students to be able to use our content in these areas, one option would be to use Apache Cordova to convert web-apps into native Android Apps with a wrapper. We attempted this briefly to see if it is possible to do so, but did not have time to implement the entire web-app into an Android app during the internship.

Additional future work includes opening up the web-app to more languages and more age groups. Our initial target audience is in India with children aged 9-10 years, but we envisage that the application can become more widely relevant over time, including socio-economically disadvantaged groups in the U.K. Note that the web-app is hosted on a public URL, and therefore is freely available to all Internet users worldwide, even though for now its intended use is in research fieldwork in India.

Acknowledgements
We are grateful to Jane Walsh and the reviewers involved in the Cambridge Language Sciences Incubator Fund. We thank Paula Buttery for her support and guidance, Jenna Renjie Zhou for collaborative work setting up the Flask server, plus Russell Moore, Sneha Sen, Louis Harris, Mansur Pasha, Kate Narciso, Seth Aycock, Electra Wallington, Anna Samuel and Alex Burchill for their valuable feedback.