A Crowdsourcing Approach to Collecting 399 Tutorial Videos on Logarithms

Jacob Whitehill
Office of Vice Provost for Advances in Learning
Harvard University
Cambridge, MA, USA
jacob_whitehill@harvard.edu

Margo Seltzer
School of Engineering and Applied Sciences
Harvard University
Cambridge, MA, USA
margo@eecs.harvard.edu

ABSTRACT

We present work on how to "teachersource" novel tutorial videos on topics related to logarithms. Specifically, we created a Human Intelligence Task (HIT) for the Amazon Mechanical Turk and collected 399 unique explanatory videos from 66 unique teachers in approximately 4 weeks. Coding of the videos is still ongoing, but initial analysis suggests significant variety of presentation format, pedagogical style, and language. In a follow-up experiment to assess the pedagogical effectiveness of the videos, we found that the best videos were statistically significantly more effective at increasing students' learning gains (posttest minus pretest) compared to a control video on a math topic unrelated to logarithms. The next step in the project is to create an intelligent decision-engine to assign tutorial videos to students based on joint properties of the video, the student, and the teacher.

1. INTRODUCTION & RELATED WORK

One of the perpetual challenges in teaching is to find good explanations that motivate and help students to grasp key concepts of the curriculum. Based on a variety of cognitive, emotional, and social factors, an explanation that is illuminating for one student may not be effective for others. On the other hand, it is often difficult – especially for a single teacher – to devise different ways of explaining concepts to satisfy the needs of all learners.

One recently proposed and promising approach to gathering data about educational resources is to crowdsource information from learners. This process, sometimes known as *learnersourcing*, has been used, for example, to identify which parts of lecture videos are confusing [2] and to describe the key instructional steps [4] and subgoals [3] of "how-to" videos. In this paper, we investigate whether an analogous process could feasibly be used to collect a wide variety of explanations from teachers, which has been dubbed *teachersourcing* [1]. The "teachers" who author the explanations could be expert teachers, ordinary people who are knowl-

edgeable on the subject matter, or even students themselves who have recently learned the material. Crowdsourcing explanations from a diverse pool of contributors can benefit students in several ways: (1) students can learn from teachers whose teaching style more closely matches their learning style; (2) some students may feel more comfortable being taught by a peer than from a professor; and (3) peer teachers may remember more vividly what the challenges were in learning the material, which can result in more effective explanations.

In contrast to recent work by [8] on crowdsourcing *text-based* explanations, here we are concerned with how to collect a diverse set of tutorial *videos* to solve specific math problems. Such a crowdsourced dataset could be used as a supplemental resource to students in traditional classrooms. It could also constitute the "action space" of an intelligent tutoring system that optimally decides which explanation to serve to given students based on characteristics of the video, the student, and the author of the explanation.

The rest of this paper presents our first experiment on crowdsourcing novel tutorial videos. As the target curriculum we chose introductory logarithms, which are sufficiently esoteric that many people – even those who once learned them many years ago – know nothing about them, but also sufficiently prosaic that many (other) people can teach them. The key questions we investigate include: Can we efficiently crowdsource novel video-based explanations (e.g., a Khan Academy video) of logarithms from sites such as the Amazon Mechanical Turk? Are the explanations mathematically correct? In what ways are they diverse? In subsequent work, we will estimate the pedagogical effectiveness of these explanations in an empirical study.

2. EXPERIMENT 1: CROWDSOURCING EXPLANATIONS

We crowdsourced novel video-based explanations of how to evaluate and solve simple logarithmic expressions and equations from "teachers" consisting of ordinary people around the world who participate in Amazon Mechanical Turk. Specifically, we created a list of 18 problems (see Figure 1) based on a previous experiment [5] on mathematics tutoring. For each problem, we solicited workers on the Mechanical Turk to produce a video (for $5 each) to explain how to solve this problem to a student. Teachers were allowed to create a different video for each problem if they desired (but not multiple videos for the same problem).

Consent Form & Video Recording Release Form

... You will then be asked to create a novel video in which you explain how to solve a short mathematical exercise: <PROBLEM>. The content and format of the video are up to you, but the video must address the problem and must be mathematically correct. For example, the video might contain a screencast showing an electronic "blackboard" on which you explain how to answer the problem. Alternatively, you might prefer to talk into a web camera and record a video of your face and your voice. ...

Survey

Please answer the questions below. When you are done, click "Next".

1. How old are you (in years)?
2. What is your gender?
3. What is the highest level of education you have completed? ...
4. How much do you enjoy mathematics? ...
5. How do you prefer to learn something new? ...

Sample Problems & Explanations

This page contains some example videos that explain how to solve math problems. Please watch the videos carefully so you know what we are looking for in this HIT.


Hints on Making a Good Video

When you make your video, you may sometimes record images of your own handwriting. Please look at the following handwriting examples so you know what distinguishes a good video from a bad video. Note that a bad video may be rejected due to poor image quality.

The following 2 examples are **OK** – the writing is dark, big, and clear.
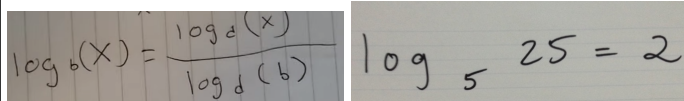
The following 3 examples are **not OK** – the writing is too small, blurry, and/or hard to read.
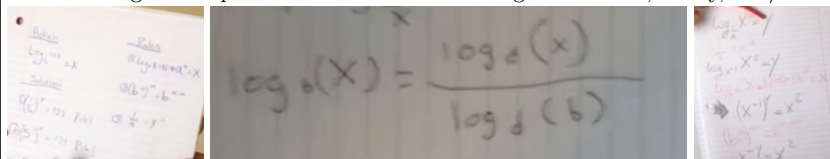
Problem & Instructions

Please examine the following math problem: <PROBLEM>

Instructions:

1. Think carefully about how you would explain to someone else how to solve this problem.
2. Create a video that explains how to solve the problem.
3. Upload the video to our server.

Rules:

- Your video must explain how to answer the following math problem: <PROBLEM>
- Your video must be original - it cannot be an existing video.
- Your video must be mathematically correct.
- Your video may not contain any images of children (<18 years old).
- Your video may not contain any nudity or profanity.

Submission

...

Figure 2: The different screens of the Human Intelligence Task (HIT) posted to Amazon Mechanical Turk to crowdsource explanations from amateur "teachers" in Experiment 1.

Basic Logarithms

Simplify:

$\log_3 1 =$ $\log_9 1 =$

$\log 100 =$ $\log_{\frac{1}{5}} 125 =$

$\log_{10} 1000 =$ $\log_{\frac{1}{x}} x^2 =$

$\log_3 81 =$ $\log_w \frac{1}{w} =$

$\log_2 8 =$ $\log_{\frac{1}{2}} \frac{1}{4} =$

Logarithms and Variables

Simplify:

$\log_a a^2 =$ $\log_x x^4 =$

$\log_4 4^{2b} =$ $\log_{x-1}(x-1)^y =$

Equations with Logarithms

Solve:

$\log_3(x-1) = 4$ $x \log_4 16 = 3$

$z \log_{10} \sqrt{10} = 4$ $y \log_{10} 1000 = 3$

Figure 1: The 18 math problems that the crowdsourced teachers were asked to solve and explain in novel video tutorials.

The Human Intelligence Task ("HIT") posted to Mechanical Turk was constructed as a "file upload" task, which means that Amazon's own servers manage the uploading and storage of data. All participants were first required to give informed consent (the experiment was approved by Harvard's IRB: IRB15-0867) and also sign a video recording release form so that their video explanations can be used in subsequent experiments on learning. We also asked teachers to complete a simple survey about their age, gender, level of education, interest in mathematics, and preferred learning style. Next, we showed several Youtube-based examples of what a good video explanation might look like. Finally, we presented the concrete problem to be solved – e.g., "Simplify $\log_2 8$" – and asked them to create and upload a video explaining how to solve it. See Figure 2 for a synopsis of the most important content of our HIT (which was rendered in HTML).

In a pilot run of the experiment, we found that several of the videos contained handwriting that was very difficult to read. We thus added explicit guidelines on handwriting quality and showed good and bad examples of each – see the "Hints on Making a Good Video" section of Figure 2. Our preliminary analysis suggests that these guidelines resulted in more legible handwriting in subsequently submitted videos.

2.1 Results

Over 2 data collection periods consisting of approximately 2 weeks each, we collected 399 videos from 66 unique teachers (17% female; minimum reported age of 18, maximum reported age of 55) that span a variety of different pedagogical approaches and presentation styles. The duration of most videos was between 1 and 3 minutes. See Figure 3 for a representative sample of the crowdsourced videos.

So far we (the authors) have personally watched 145 of the 399 submitted videos and judged them for correctness (review of the remaining videos is still ongoing). Our impression was that, while several of them contained mistakes and
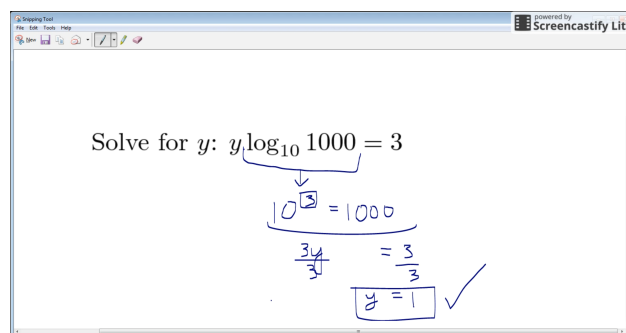

Figure 4: One of the crowdsourced explanations that uses a recorded screencast to explain how to solve the equation $y \log_{10} 1000 = 3$.

some were rather unclear, there were also many that could be very effective for helping students learn about logarithms. Below is the audio transcript of one of the explanations of how to solve the equation $y \log_{10} 1000 = 3$ (see also Figure 4):

> Solve for y in the equation: y times log base 10 of 1000 equals 3. Well, log base 10 of 1000 – this is just a constant. They're asking us for: what is the exponent we raise 10 to, to get 1000? What is the exponent we should put on 10 to get 1000? Well, 10 to the third power is 1000. 10 squared is 100; 10 cubed is 1000. So this whole expression simplifies to just 3. The logarithm means: what is the exponent? The exponent is 3. So we have y times that number, or in other word, 3 y equals 3. One more step: to get y by itself is to divide by sides by its coefficient. Divide both sides by 3. So y is equal to 1. And that's our final answer.

This explanation is well structured and suggests careful thought by the teacher on how to explain the solution. Moreover, by combining speech with video – e.g., to draw a box into which the exponent is written, and to show each line of the derivation – the explanation is arguably clearer than what a purely textual or audio-based explanation could provide.

Conducting an experiment to measure quantitatively the learning gains of the crowdsourced videos is the subject of Experiment 2 (see section below).

2.2 Correctness

We deemed a video to be "correct" if it began with the problem statement, ended with the correct solution, and contained no statement that was objectively false. Importantly, we made no attempt to assess the quality of the pedagogy – we will pursue this more rigorously and objectively in follow-up learning experiments (see Conclusions section).

117 videos were judged to be mathematically fully correct. 16 videos were judged as incorrect. Some of the mistakes were incorrect verbal usage of terminology even if the written derivation itself was correct. For example, one teacher
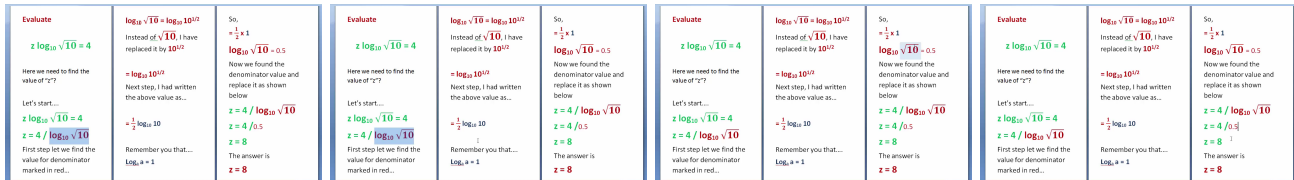
Video 1
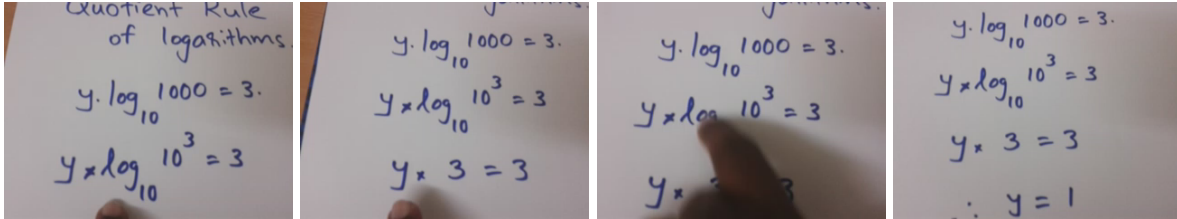
Video 2
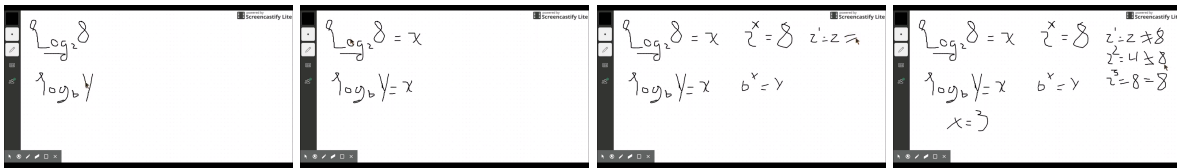
Video 3

Video 4

Video 5

Figure 3: Snapshots (at 25%, 50%, 75%, and 100% time duration within each video) of 6 representative examples of 399 total crowdsourced explanatory videos on logarithms.

read the expression $\log x$ as "log times x" instead of "log of x". Other mistakes were more egregious. For instance, in one video, the teacher "canceled" two occurrences of the log function – one in the numerator and one in the denominator:

$$\frac{\log \frac{1}{4}}{\log \frac{1}{2}} = \frac{\cancel{\log} \frac{1}{4}}{\cancel{\log} \frac{1}{2}} = \frac{\frac{1}{4}}{\frac{1}{2}}$$

(Interestingly, his final answer to the problem – due to another mistake – was actually correct.)

7 videos were considered to be "borderline". For example, some teachers referred to a mathematical *expression* (e.g., $\log_2 1$) as an *equation* even though there is no equals sign.

2.3 Format
As shown in Figure 3, there was diversity in the presentation formats and styles used in the videos. The five videos shown in the figure illustrate the most common styles; these include (1) writing on paper; (2) recording a video of the teacher's computer screen; (3) speaking directly to the learner in a face video along with written materials to show the derivation; (4) a step-by-step "Powerpoint"-style presentation; and (5) a static Powerpoint slide to which the instructor points using the mouse. Regardless of style, all videos included accompanying audio to explain the solution. Some teachers also mixed styles by writing on the Powerpoint slide.

2.4 Pedagogical style
We observed two general approaches that teachers used to derive the solutions to the problems. In some explanations, the *definition of logarithm* – i.e., the logarithm of x base b is the power to which b must be raised to equal x – was invoked to solve the problem. For example, to reduce $\log_{10} 1000$, one can use the fact that clearly $10^3 = 1000$ to arrive at the correct answer of 3. In other explanations, the teacher emphasized the *syntax* of logarithms and how rules can be applied to transform a problem step-by-step into the solution. For example, to simplify $\log_x x^4$, the teacher would note that $\log_x y^c = c \log_y x$ for all c to derive $4 \log_x x$; then, he/she would note that $\log_x x = 1$ for all x to derive $4 \times 1 = 4$.

2.5 Language
Although all crowdsourced videos were in English, there was variability in the geographical origin and dialect of the spoken English. In particular, several teachers used terminology such as "5 into x" to express the *multiplication* of 5 with x, i.e., $5x$. Although we (as American English speakers) were initially confused by this phrasing, this terminology is correct and widely used in India [6]. This also highlights the need for both a large, diverse set of explanations as well as smart decision-making in determining which learners are assigned to which explanation.

2.6 Enjoyment
Several of the workers who provided explanations in our study expressed to us (in an email) their enjoyment in completing the HIT, and many of them created explanations for several different problems. This suggests that crowdsourcing may provide a *scalable* way of collecting educational content for a variety of learning tasks.

3. EXPERIMENT 2: MEASURING LEARNING GAINS
To estimate the pedagogical effectiveness of the crowdsourced videos, we conducted a second experiment on Mechanical Turk involving 200 participants (with 0.40 reward for participation). In particular, we randomly selected 40 videos from the 145 that we had personally verified for mathematical correctness. (The number of required subjects to obtain a statistically significant result would have been very large had we used *all* of the correct videos.) In the experiment, each participant first took a pretest on logarithms that included exactly the problems shown in Figure 1. After the pretest, each subject was randomly assigned a video to watch. With probability 0.2, the assigned video was a mathematics tutorial on a topic unrelated to logarithms (specifically, on the number π – see `https://www.youtube.com/embed/7Mz7xU3zZvk`). This video serves as a "control" for the experiment. With uniform probability of $0.8/40 = 0.02$, the subject was assigned to watch one of the 40 preselected videos. After watching the video, the subject then took a posttest whose content was comparable in length and content to the pretest but contained different problems. The dependent variable was the posttest score minus the pretest score. Note that, since some subjects started but did not complete the experiment, the number of subjects collected per video varied.

Because this study is about crowdsourcing novel explanations from ordinary people around the world who may have varying mathematical skill and pedagogical expertise, we do not expect *all* the videos to be effective in helping students learn logarithms. Rather, our hope is that *some*, or even a *few* of the videos, are highly effective. In our experiment we therefore investigated whether the average learning gains ("LG" – defined as posttest minus pretest) achieved by the top K videos was higher than what we would expect by chance. Specifically, the null hypothesis H_0 was that the learning gains for any participant for any video is a normally distributed random variable with mean μ and variance σ^2, where μ is the sample mean for the control video and σ^2 is the sample variance over *all* videos and all students. We then used numerical simulation (with 5000 simulation runs for each K) to compute the probability (p-value) that the expected learning gains of the top K videos (using the same number of samples for each video as in the actual experiment) under H_0 was greater than or equal to the sample average learning gains of the actual experiment. Note that we expect the expected learning gains of the *best K* videos to be larger than 0, even under the null hypothesis.

3.1 Results
The top K crowdsourced videos were statistically significantly more effective, in terms of the learning gains (posttest minus pretest score), than what one would expect by chance, for all $K \in \{1, 2, \ldots, 10\}$. In particular, the top 3 videos (out of 40 randomly chosen videos from the 145 verified video set) helped students to improve their test score by 34%; this is statistically significantly higher than the 24% that we would expect under the null hypothesis. See Table 1 for complete results, including the average learning gains, expected learning gains given H_0, and p-value, for each K.

K	Average LG	Expected LG for H_0	p-value
1	0.47	0.28	0.02
2	0.39	0.25	0.02
3	0.34	0.23	0.02
4	0.31	0.22	0.02
5	0.29	0.21	0.02
6	0.27	0.20	0.02
7	0.26	0.19	0.02
8	0.24	0.18	0.02
9	0.23	0.18	0.01
10	0.23	0.17	0.01

Table 1: Results of Experiment 2 showing the effectiveness, in terms of learning gains ("LG" – defined as posttest minus pretest score) of the top K most effective videos, for $K \in \{1, \ldots, 10\}$. The second column shows the average learning gains observed empirically. The third column shows the expected learning gains under the null hypothesis H_0. The fourth column shows the p-value.

4. CONCLUSIONS

We have presented research on how to "teachersource" a large variety of video-based explanations on topics related to logarithms. Specifically, we described the Human Intelligence Task (HIT) we created for Amazon Mechanical Turk to collect 399 novel videos from 66 unique teachers. Our conclusion from this work is that crowdsourcing of educational videos from ordinary people is feasible – provided that appropriate guidelines on how to craft the explanations are given to the workers. The pedagogical effectiveness of the *best* videos, as expressed in the learning gains (posttest minus pretest scores) achieved by students who watch those videos, was statistically significantly higher than what we would expect from watching mathematics videos on topics unrelated to logarithms.

In addition, we observed considerable diversity of the tutorial videos we solicited in terms of presentation format, pedagogical style, and language, as well as the specific phrasing of the explanation itself. The findings of our work also indicate that – as with all crowdsourcing tasks – it is important to implement sufficient quality-control procedures before offering the explanations to real students.

Future work: Given empirical results about which particular crowdsourced videos were more effective than others (Experiment 2), we can run a more targeted experiment to estimate the learning gains for the most effective videos. In addition, we will also investigate machine learning-based methods (e.g., [8, 7]) to determine which students should receive which explanations based on joint properties of students and teachers.

Acknowledgement

5. REFERENCES

[1] N. T. Heffernan, K. S. Ostrow, K. Kelly, D. Selent, E. G. Inwegen, X. Xiong, and J. J. Williams. The future of adaptive learning: Does the crowd hold the key? *International Journal of Artificial Intelligence in Education*, pages 1–30, 2016.

[2] J. Kim, P. J. Guo, D. T. Seaton, P. Mitros, K. Z. Gajos, and R. C. Miller. Understanding in-video dropouts and interaction peaks inonline lecture videos. In *Proceedings of the first ACM conference on Learning@ scale conference*, pages 31–40. ACM, 2014.

[3] J. Kim, R. C. Miller, and K. Z. Gajos. Learnersourcing subgoal labeling to support learning from how-to videos. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pages 685–690. ACM, 2013.

[4] J. Kim, P. T. Nguyen, S. Weir, P. J. Guo, R. C. Miller, and K. Z. Gajos. Crowdsourcing step-by-step information extraction to enhance existing how-to videos. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 4017–4026. ACM, 2014.

[5] L. P. Salamanca, A. R. Carini, M. A. Lee, K. Dykstra, J. Whitehill, D. Angus, J. Wiles, J. S. Reilly, and M. S. Bartlett. Characterizing the temporal dynamics of student-teacher discourse. In *Development and Learning and Epigenetic Robotics (ICDL), 2012 IEEE International Conference on*, pages 1–2. IEEE, 2012.

[6] Wikipedia. into, 2016. https://en.wiktionary.org/wiki/into; [Online; accessed 4-March-2016].

[7] J. J. Williams, N. Ali, J. Kim, J. Whitehill, S. Maldonado, M. Pechnizkiy, L. Chu, and N. Heffernan. The mooclet framework: Improving online education through experimentation and personalizaton of modules. *SSRN*, 2014.

[8] J. J. Williams, J. Kim, A. Rafferty, S. Maldonado, W. Lasecki, K. Gajos, and N. Heffernan. Axis: Generating explanations at scale with learnersourcing and machine learning. In *ACM Learning at Scale*, 2016.