Measuring the Video-driven Discussion Engagement Using IRT model

Yi Chen, Patricia Alejandra Behar, Xiang Liu, HuiSoo Chae, Gary Natriello
EdLab, Teachers College, Columbia University

Abstract

In this study, we introduced a general methodology, three parameter Poisson (3PP) model, for online video-driven discussion (VDD) engagement measurement. To make the model more robust and apply the model in small sample size case, we improved Duplicate, Erase and Replace (DupER) Augmentation in 3PP model. Finally, we conducted behavioral analyses to gain insights into patterns of students comment interaction on a video-driven discussion platform, Vialogues, in a real class environment.

Keywords: three parameter Poisson model, DupER Augmentation, video-driven dicussion, engagement measurement

1. Introduction

Video-driven discussion (VDD) has been widely used for decades in education to promote reflection, critical thinking, and constructive learning (Copeland & Decker, 1996; Koc, Peker, & Osmanoglu, 2009; Close, Scherr, Close, & McKagan, 2012). While traditional video learning materials alone support learning passively, video-driven discussion platforms provide an active learning environment through asynchronous discussion and content sharing (Sherin, 2003b).

With the development of innovative technologies, a growing number of online video-discussion learning tools have been created and widely applied in education (Giannakos, Chorianopoulos, & Chrisochoides, 2015). Examples including YouTube (http://youtube.come), Vimeo (http://vimeo.com), TED-Ed (http://ed.ted.com), and Vialogues (https://vialogues.com). However, the majority of these online resources have not been consistently applied in a real class environment. Furthermore, the increasing volume of user data has not been explored fully to understand the students' learning behavior.

Item response theory (IRT) offers flexible and useful models for assessment data. Using a similar idea, people could apply IRT models in measuring participants' engagement level and item popularity in a VDD environment. However, their use is limited due to the need for a large sample size. This, to some extent, explain why IRT models have not been widely used in a real class environment where enrollment is typically less than 100.

In this study, we introduce a general approach, 3 parameter Poisson (3PP) model, for measuring the online discussion behavior. This model combines the idea of IRT models and the framework of the zero-inflated Poisson (ZIP) model. In addition, we improve and apply Duplicate, Erase and Replace (DupER) Augmentation to overcome the large sample size requirement of the IRT model. Finally, we use our model to analyze real data as an empirical example.

2. Literature Review

Video-driven discussion in Education. Using Video-driven discussion technique in education has been widely studied by researchers (de la Torre & Hong, 2010; Poquet, Lim, & Dawson, 2018; Schwan & Riempp, 2004; Vieira, Lopes, & Soares, 2014; Lee & Sharma, 2008; Traphagan, Kucsera, & Kishi, 2010; Ljubojevic, Vaskovic, Stankovic, & Vaskovic, 2014). According to this research, the benefits of VDD in education can be summarized into four categories. Firstly, VDD provides a new opportunity to manipulate and interpret the principles and processes situated in the video. Secondly, VDD links to content and concepts to everyday experience. Thirdly, VDD provides a tool for self-evaluation, modifying, testing and revising one's knowledge. Fourth, VDD could inspire and engage students.

An increasing number of the educational researchers and practitioners have considered the application of video-discussion in both traditional teaching environments and in distance education system (Haga, 2002; Teng & Taveras, 2004; Ljubojevic et al., 2014; Lord, 1983; Sherin, 2003a). With the development of VDD techniques and their application in digital platforms, many researchers have begun to study student behavioral patterns in various e-learning environments. For example, (Brooks, Epp, Logan, & Greer, 2011) explored students' engagement patterns with video lectures.Mirriahi, Liaqat, Dawson, and Gašević (2016) analyzed user profiles for video annotation tools. Our current work seeks to build upon this evolving body of research.

Small Sample Size IRT models. The item response theory (IRT) model is a well-known and widely accepted model that is used across a variety of assessment programs. The primary benefit of IRT is that estimates of item parameters are sample independent, and ability estimates are independent of items (Stenbeck, Hambleton, Swaminathan, & Rogers, 1992). However, IRT also requires a relatively large sample size to obtain accurate parameter estimations. For example, three parameters logistic model (3PL) empirically requires about 60 items and 1000 examinees to obtain adequate estimates (Swaminathan & Gifford, 1986). This limits the application limitations of IRT model in the real class environments where class size are relatively small.

Generally speaking, there are four approaches to obtain accurate parameter estimates with small sample size IRT models (Hula, Fergadiotis, & Martin, 2012). The first approach is to use modified IRT models. Typically, investigator place restrictions on one or more parameters. An example of this is fixing lower asymptotes in 4 parameter model (4PL) to a common non-zero value for all items (Wise, 1991). Harwell and Janosky (1991) and Roberts (2008) also proposed modified IRT models under small sample size.

The second approach is optimal sampling or estimation techniques. For example, McNeish (2016) used a Bayesian estimator was used to account for the small sample size, providing a more trustworthy result than a traditional maximum likelihood estimator.

The third approach is adding more auxiliary data information in parameter estimation. An example might be incorporating examinees and items' background information. This data is assumed to provide extra information and consequently lead to more accurate estimation (Forero & Maydeu-Olivares, 2009).

The last approach is to use a resampling technique. The two most common technique is jackknifing and bootstrapping (He, 2017). DupER Augmentation is another approach in the family of nonparametric statistical techniques. Different from jackknifing and bootstrapping which are aimed at deriving a robust estimation of standard errors and confidence intervals of population parameters, DupER is used for creating a pseudo data set and parameter estimates.

3. Methodology

Three Parameter Poisson Model. In this study, we introduced a three parameter Poisson (3PP) model for the real class online discussion environment, Vialogues. 3PP is a combination of zero-inflated Poisson (ZIP) model and a 3PL model. We define y_{ij} as the comment word count for the i-the student on the jth vialogue (i.e., video discussion). Here we define the three parameters in 3PP and explain their real meanings.

Table 1
Model Specification

Parameter	Symbol	number	real meaning
participant parameter	θ_i	I	engagement level for i-th participant
item parameter	β_j	J	popularity level for j-th item
course parameter	ϕ	1	active level for the course

The basic assumptions are: 1) The bigger value of a θ leads to the more expected word count ($E(y_{ij})$) and represents a higher engagement level of the participant; 2) The bigger value of a β leads to the more expected word count ($E(y_{ij})$) and represents a higher popularity level of the item; 3) The bigger value of a ϕ leads to an expected word count y_{ij} bigger than zero (participant posts at least a one-word comment) and represents the higher activity level of the whole course.

The combination of $\theta_i + \beta_j$ is the rate parameter in the Poisson distribution. In general, the baseline model can be expressed as follows:

$$P(y_{ij}|\theta_i, \beta_j) = \frac{(\theta_i + \beta_j)^{y_{ij}} e^{(\theta_i+\beta_j)}}{y_{ij}!} \tag{1}$$

Overdispersion is a common phenomenon in the analysis of discrete data. In particular, we do not expect that a participant would post commons on all items. In other words, a certain number of item response y_{ij} would be 0. A common solution to this problem is using a zero-inflated model. Course parameter ϕ, in the following function, actually represents the probability of item response expected to be bigger than zero.

$$P(y_{ij}|\phi, \theta, \beta) = \begin{cases} (1-\phi) + \phi \times Poisson(y_{ij}|\theta_i + \beta_j) & y_{ij} = 0 \\ \phi \times Poisson(y_{ij}|\theta_i + \beta_j) & y_{ij} > 0 \end{cases} \tag{2}$$

In the 3PP model the mean and variance for every observation are: $\mu = E(y_{ij}) = (\theta_i + \beta_j)(1 - \phi)$ and $Var(y_{ij}) = \mu + (\frac{\phi}{1-\phi})\mu^2$

DupER Augmentation. Duplicated, Erase and Replace (DupER) Augmentation is a new sample size augmentation technique. DupER serves as a means to generate additional plausible observation to augment the raw data set. This method works as follows:

1. **Duplicate**: Given a data set of the participant, their response vector is duplicated several times. For example, for a 30 student class with 10 items, we can duplicate their response vector 10 times. Now, we have a new dataset with 30 raw data and 300 duplicated response vectors;

2. **Delete**: For the observation in the new data set, sample randomly sampling a certain proportion of observation from the duplicated data with replacement. Then, delete these observations. In this way, we can make the new data set missing at random (MAR). For example, for the 300 duplicated response vector, we have $300 \times 10 = 3000$ observations. We randomly sample 40% observations with replacement and set them as missing data.

3. **Replace** The missing observations are replaced using imputation with the whole data set. For example, after deleting random observations, we combine the raw data set with the duplicated data set which is a 330×10 matrix. Then we could use imputation techniques to generate plausible values.

A simple illustration of DupER Augmentation is shown in Figure 1. DupER assumes the raw data set is representative of the population and variation of the population can be realized by deleting observations at random and imputation. Since all the direct information is coming from the raw data, this method has the risk of overfitting.

Imputation involves using the available data to determine plausible values for missing data. In general, there are two types of imputation, single imputation, and multiple imputations. Predictive mean matching, linear regression imputation, stochastic regression imputation, and EM imputation are the most widely used single imputations. Multiple imputations expand single imputation by creating multiple imputed data sets. The typical one is the Markov chain Monte Carlo (MCMC) imputation. The choice of the imputation method could influence the performance of DupER, however, there is no agreement as to which method is the best under the different study condition. In our study, we use MCMC imputation. The reason for using MCMC imputation lies in its stochastic properties (Schunk, 2008). In other words, imputed values are selected randomly from a distribution: two response vectors with the same missing data pattern may result in different response patterns after imputation.

When this technique first was introduced by Brette Patrick Foley, he permitted deletion and replacement of the observations in the original data set in the second and third steps. We avoid this approach since we want to keep the original item response pattern and consequently improve the measurements of participant engagement for the original data set. Moreover, we only care about the participant parameters for the raw data set.

Bayesian Approach for Parameter Estimation. In this study, we took a Bayesian approach to model estimation. Generally, Bayesian statistics use the MCMC algorithm for estimating the posterior distribution. Investigators typically transform parameter space through exponentiation since this would speed up the estimation.

One of the most important steps in Bayesian statistics is to choose the prior distributions. When the prior is set for matching an expert's belief about the parameters, Raw Response Patterns

	1	2	3	4	5
A	24	13	11	31	0
B	0	8	10	0	3

Raw Response Patterns & Duplicated Response Patterns

	1	2	3	4	5
A1	24	13	11	31	0
B1	0	8	10	0	3
A2	24	13	11	31	0
B2	0	8	10	0	3
A3	24	13	11	31	0
B3	0	8	10	0	3

Randomly Delete Observation in Duplicated Response Patterns (40%)

	1	2	3	4	5
A1	24	13	11	31	0
B1	0	8	10	0	3
A2	NA	13	11	NA	NA
B2	0	8	NA	0	3
A3	24	13	11	31	0
B3	NA	NA	10	NA	NA

Impute Missing Values

	1	2	3	4	5
A1	24	13	11	31	0
B1	0	8	10	0	3
A2	17	13	11	27	8
B2	0	8	2	0	3
A3	24	13	11	31	0
B3	3	5	10	0	12

Figure 1. Illustration of DupER for 10 items and 2 participants

it's called the subjective prior. Otherwise, investigators use objective prior. Examples are the default, non-informative and low-informative prior that essentially reflects a lack of strong and precisely quantified prior information. For θ, we set low-informative priors: $log(\theta) \sim Cauchy(0, 1)$. In this way, we fix the scale of participant parameters and ensure the identification of the model. For item parameters, we set $log(\beta)$ an improper prior as a uniform distribution. Since the parameter ϕ has the range from 0 to 1, the most common noninformative prior is to use the uniform distribution. $\phi \sim Uniform(0, 1)$

4. Empirical Example

The present study's primary data is collected from the real class environment at a Graduate School of Education. In this class, a variety of online learning platforms are encouraged to be used for course discussions and team projects. One of the web application is called Vialogues (Agarwala, Hsiao, Chae, & Natriello, 2012), which is a video-driven discussion tool developed by EdLab at Teachers College. A vialogue consists of a video and all the discussions associated with the video.

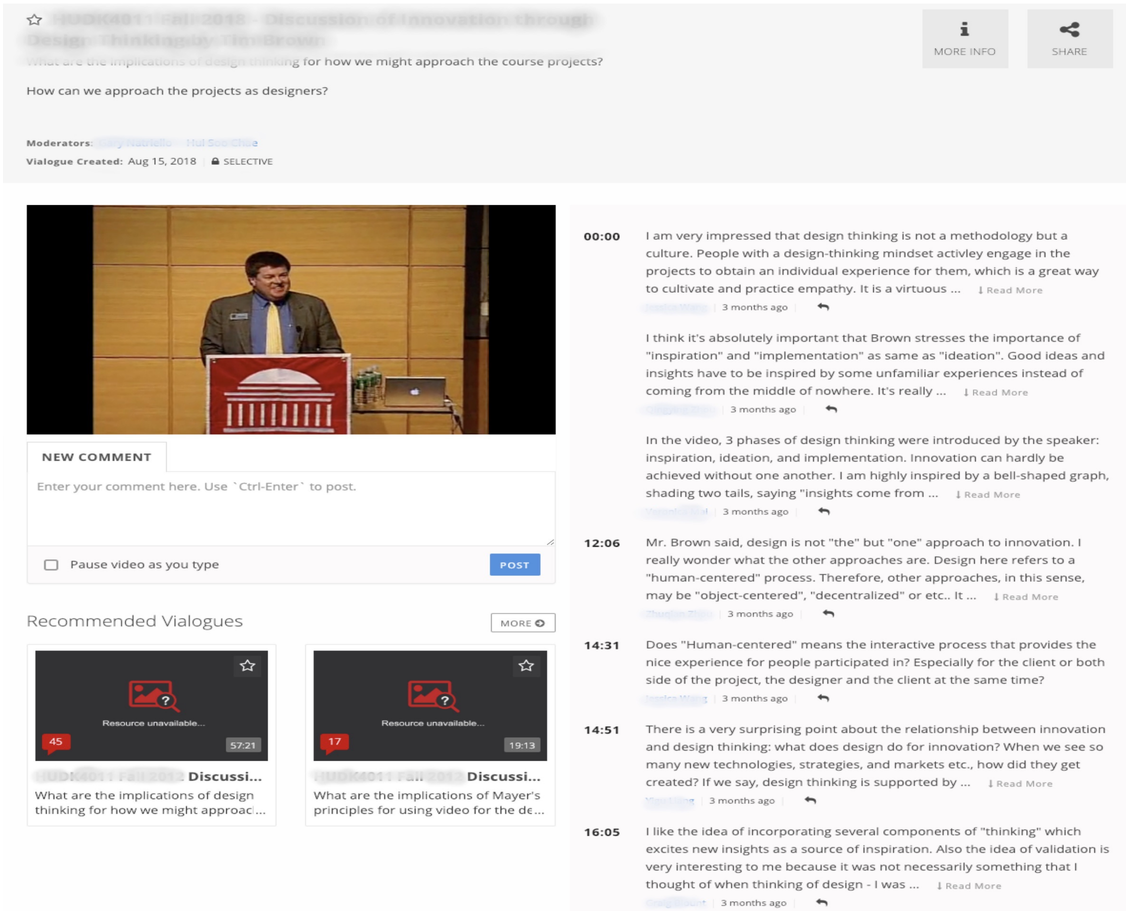

Figure 2. Vialogues User Interface

Data. Throughout the semester, 28 vialogues have been actively discussed by the students. Among these vialogues, 14 videoes were uploaded by instructors as course resources and 14 videoes were uploaded by students for team project presentations. All 28 students involved in this class were encouraged to post new comments or make replies freely. The length of each post is not limited. In addition, a reply to a comment is viewed as a comment as well. Instead of focusing on the number of comments, we use the real word count of student posts for each vialogue. Here is an example of a conversation between two students: **Student 1**: *I think a lot after watching this video. As Jason says, "if algorithms are going to curate the world for us, we need to make sure that they are not just keyed to relevance..." Because the internet helps us, the people from all over the world, to connect and to know the world, software engineers should decide what searchers get to see more carefully. Relevance is important for websites like shopping websites, however, it might not be extremely important for other kinds of websites. When it comes to learning, my high school teacher told me that we should read as many books as we can, which are not restricted to the academic perspective (i.e., relevant). If we get enough "input", we can "output" unconsciously.* **Student 2 (reply)**: *Interesting, I aggree.* In this example, both students post a single comment, while the information they provide can be different. Treating every

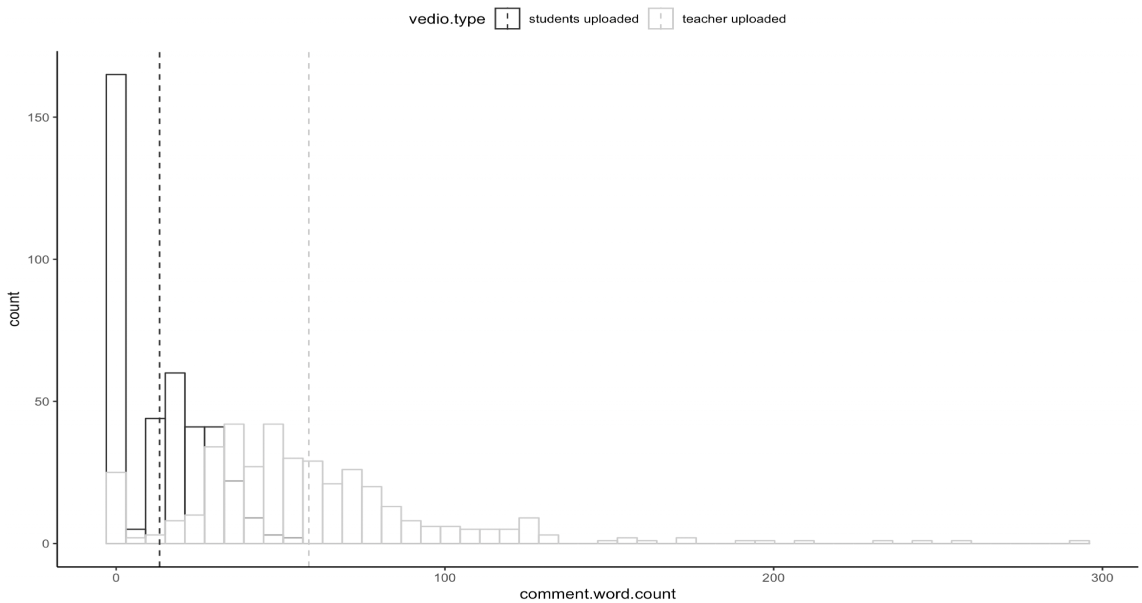

Figure 3. Comment Count Distribution

comment equally brings unignorable bias in engagement measurement. Through counting words, we give more weight to the long comments. In this way, we distinguish the deep and comprehensive interactions from the short and common ones.

To reduce the noise from original text content, we conducted the following text data processing operations: word segmentation, removing punctuation, deleting of numbers, updating the stopwords, removing stopwords, and deleting white space. As we can see from figure 3, the videos that are uploaded by the teacher a higher expected word count and bigger variance than those uploaded by students. On the other hand, there exists a clear mixture of distribution patterns. Overall, nearly one-fourth of the cases, the word counts are 0. In other words, in 24 of 100 times on average, the student would not make any real comments on the VDD platform. This value gives us a big picture of the activity level in this class.

Model Check with Parameter Recoverage. In this section, we simulate the value of the parameter from the prior distribution. The, based on our model assumption, we generate fake data with the known parameter value. Using our model, we can get the estimates with the fake data. By comparing the known true parameter value with our estimates, we can a general sense of the usefulness of our model.

As we can see in Figure 4, the posterior predictive distribution for all parameters covers the true value of parameters with reasonable probability. This result, to some extent, proves our model is valid.

Posterior Prediction Check. In the words, the great statistician George Box mentioned: "models are an approximation." With these words in mind, posterior predictive check (PPC) is an important tool in Bayesian statistics to validate a model (Kruschke, 2013). Elaborating slightly, PPCs analyze the degree to which data generated from the model deviate from data generated from the true distribution.

According to Figure 5, we can say that the expected common word count match with

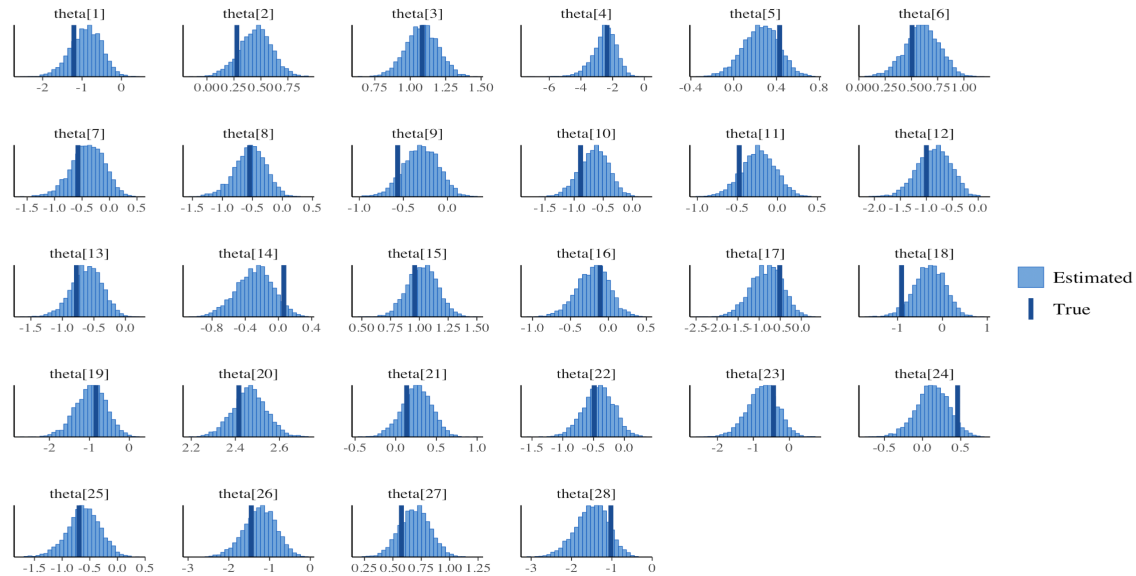

Figure 4. Parameter Recoverage Check

the real data. Thus, our model result is reliable.

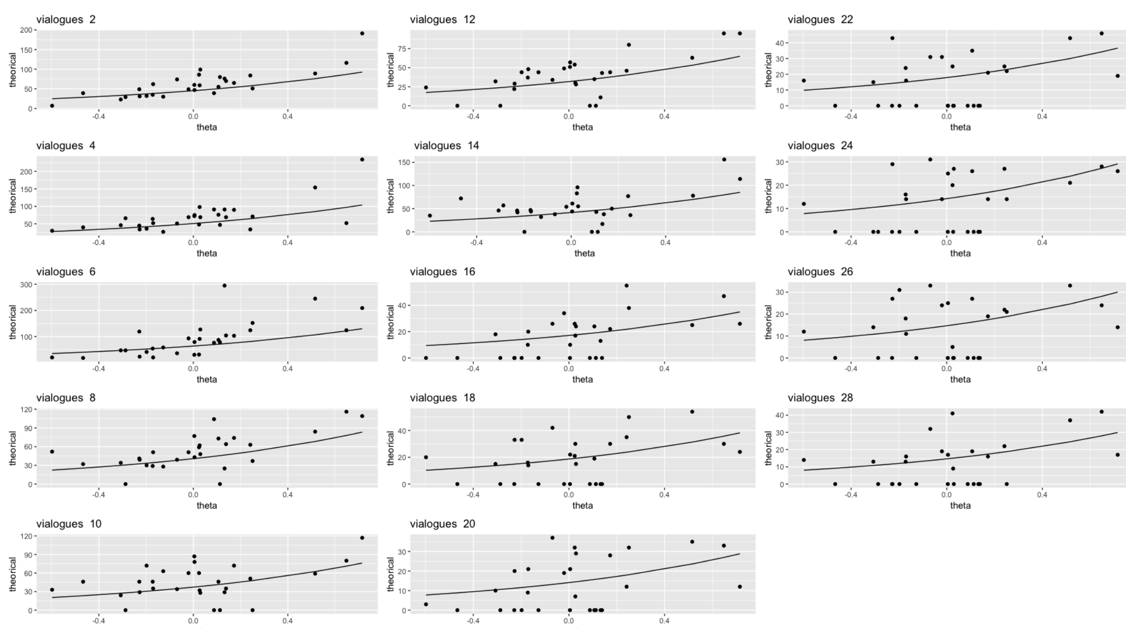

Figure 5. Posterior prediction check for real data

Results & Interpretation. In this study, we use the R (Team & R Development Core Team, 2016) and stan (Ben-Yehuda, 2013) for model estimation. As we can see from the model results in table 2, R-hat statistics for every parameter are 1. R-hat quantifies the consistency of an ensemble of Markov chains. In general, we require R-hat to be smaller than 1.1. At the same time, the standard deviation is small (all less than 0.1) which prove that we get a more accurate estimation.

The estimated value of β. As we can see from the item parameters, the first 14 parameters have higher values than the last 14 parameters. This indicates that the first 14 vialogues are more popular. This result is reasonable since the first 14 values are uploaded by instructors while the last 14 vialogues are uploaded by students themselves. We can expect that students would pay more attention to the items created by instructors and post more comments. At the same time, ϕ is estimated to be between 23% and 25%. This means 23 to 25 of 100 times, the students would not make any real comment when they're watching the video on the vialogue. As we can see from figure 6, the scale of the participant parameters is fixed to 1. Student 2, 15 and 25 are the three most active students. The posterior median of their latent parameter θ is 2.03, 1.91 and 1.67. Similarly, student 1 and 12 are the least active students. But we cannot tell whether student 12 is indeed more active than student 1 since the 95% confidence intervals overlap.

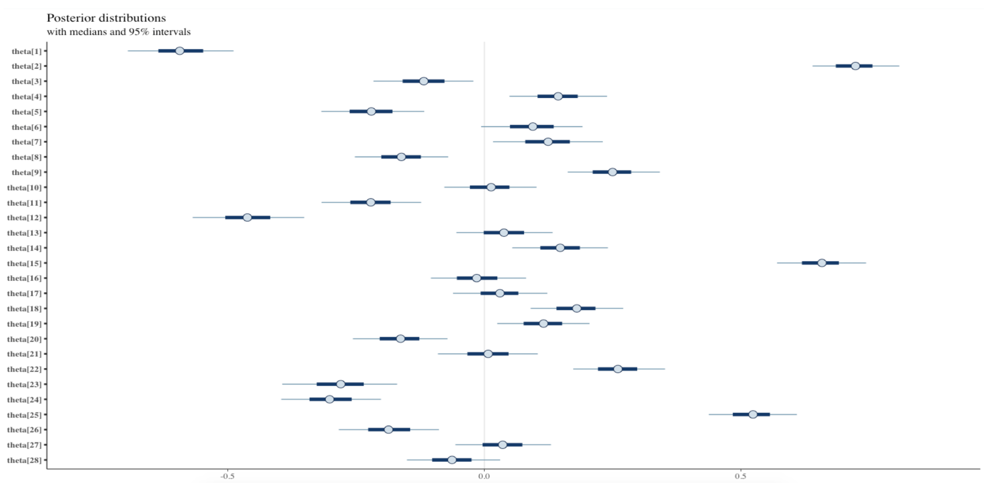

Figure 6. Comment Count Distribution

5. Discussion

Video-driven discussion plays an increasingly important role both for the distance education system and in traditional classroom environments today. Although, many researchers have discussed the benefits and limitations of using VDD in education, the growing volume of students' data in video discussion system has not been deeply explored, perhaps because we lack a quantitative methodology for measuring participants' behavior on video-driven discussion systems, particularly when they are used in typical classes. Our study provides a general solution to this problem. The 3PP model enables researchers to measure the participants' engagement levels and the popularity levels of the video material. At the same time, the estimation of parameters is sample independent. The third parameter ϕ, on the other hand, provides an overall measurement of activity level for the whole course. Moreover, DupER Augmentation makes the application of the 3PP model in a small size class environment possible.

However, there still have many challenges related to this topic. Firstly, comment word count as the unit of observation seems to be more reasonable than simple comment count. But a longer comment does not necessarily represent more enthusiasm and deeper understanding. Second, our study does not pay enough attention to the long-term learning process. In other words, the engagement level and popularity level could change over time. Finally, 3PP provides a framework to analyze the hidden factor which may influence participants' discussion behavior. In particular, using a similar idea of Differential Item Functioning (DIF), we could identify the potential source of bias in person measurement in the future.

Acknowledges

We would like to thank EdLab at Columbia University for their support for this study.

References

Agarwala, M., Hsiao, I. H., Chae, H. S., & Natriello, G. (2012). Vialogues: Videos and dialogues based social learning environment. In *Proceedings of the 12th ieee international conference on advanced learning technologies, icalt 2012.* doi: 10.1109/ICALT.2012.127

Ben-Yehuda, N. (2013). *Stan.* doi: 10.1177/1741659013488472

Brooks, C., Epp, C. D., Logan, G., & Greer, J. (2011). The who, what, when, and why of lecture capture. In *Proceedings of the 1st international conference on learning analytics and knowledge - lak '11.* doi: 10.1145/2090116.2090128

Close, E. W., Scherr, R. E., Close, H. G., & McKagan, S. B. (2012). Development of proximal formative assessment skills in video-based teacher professional development. In *Aip conference proceedings.* doi: 10.1063/1.3679983

Copeland, W. D., & Decker, D. L. (1996). Video cases and the development of meaning making in preservice teachers. *Teaching and Teacher Education*. doi: 10.1016/0742-051X(95)00058-R

de la Torre, J., & Hong, Y. (2010). Parameter estimation with small sample size a higher-order IRT model approach. *Applied Psychological Measurement*. doi: 10.1177/0146621608329501

Forero, C. G., & Maydeu-Olivares, A. (2009). Estimation of IRT Graded Response Models: Limited Versus Full Information Methods. *Psychological Methods*. doi: 10.1037/a0015825

Giannakos, M. N., Chorianopoulos, K., & Chrisochoides, N. (2015). Making sense of video analytics: Lessons learned from clickstream interactions, attitudes, and learning outcome in a video-assisted course. *International Review of Research in Open and Distance Learning*. doi: 10.19173/irrodl.v16i1.1976

Haga, H. (2002). Combining video and bulletin board systems in distance education systems. *Internet and Higher Education*. doi: 10.1016/S1096-7516(02)00084-2

Harwell, M. R., & Janosky, J. E. (1991). An Empirical Study of the Effects of Small Datasets and Varying Prior Variances on Item Parameter Estimation in BILOG. *Applied Psychological Measurement*. doi: 10.1177/014662169101500308

He, X. (2017). Resampling methods. In *Handbook of quantile regression.* doi: 10.1201/9781315120256

Hula, W. D., Fergadiotis, G., & Martin, N. (2012). Model choice and sample size in item response theory analysis of aphasia tests. *American Journal of Speech-Language Pathology*. doi: 10.1044/1058-0360(2011/11-0090)

Koc, Y., Peker, D., & Osmanoglu, A. (2009). Supporting teacher professional development through online video case study discussions: An assemblage of preservice and inservice teachers and the case teacher. *Teaching and Teacher Education*. doi: 10.1016/j.tate.2009.02.020

Kruschke, J. K. (2013). *Posterior predictive checks can and should be Bayesian: Comment on Gelman and Shalizi, 'Philosophy and the practice of Bayesian statistics'.* doi: 10.1111/j.2044-8317.2012.02063.x

Lee, K. J., & Sharma, M. D. (2008). Incorporating active learning with videos: A case study from physics. *Teaching Science - The Journal of the Australian Science Teachers Association*.

Ljubojevic, M., Vaskovic, V., Stankovic, S., & Vaskovic, J. (2014). Using supplementary video in multimedia instruction as a teaching tool to increase efficiency of learning and quality of experience. *International Review of Research in Open and Distance Learning*. doi: http://dx.doi.org/10.19173/irrodl.v15i3.1825

Lord, F. M. (1983). Statistical bias in maximum likelihood estimators of item parameters. *Psychometrika*. doi: 10.1007/BF02293684

McNeish, D. (2016). On Using Bayesian Methods to Address Small Sample Problems. *Structural Equation Modeling*. doi: 10.1080/10705511.2016.1186549

Mirriahi, N., Liaqat, D., Dawson, S., & Gašević, D. (2016). Uncovering student learning profiles with a video annotation tool: reflective learning with and without instructional norms. *Educational Technology Research and Development*. doi: 10.1007/s11423-016-9449-2

Poquet, O., Lim, L., & Dawson, S. (2018). Video and Learning : A Systematic Review (2007-2017). *Proceedings of the International Conference on Learning Analytics and Knowledge*. doi: 10.1145/3170358.3170376

Roberts, J. S. (2008). Modified likelihood-based item fit statistics for the generalized graded unfolding model. *Applied Psychological Measurement*. doi: 10.1177/0146621607301278

Schunk, D. (2008). A Markov chain Monte Carlo algorithm for multiple imputation in large surveys. *AStA Advances in Statistical Analysis*. doi: 10.1007/s10182-008-0053-6

Schwan, S., & Riempp, R. (2004). The cognitive benefits of interactive videos: Learning to tie nautical knots. *Learning and Instruction*. doi: 10.1016/j.learninstruc.2004.06.005

Sherin, M. G. (2003a). *1. NEW PERSPECTIVES ON THE ROLE OF VIDEO IN TEACHER EDUCATION.* doi: 10.1016/S1479-3687(03)10001-6

Sherin, M. G. (2003b). Using Video in Teacher Education. *Advances in Research on Teaching*. doi: 10.1016/S1479-3687(03)10001-6

Stenbeck, M., Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1992). Fundamentals of Item Response Theory. *Contemporary Sociology*. doi: 10.2307/2075521

Swaminathan, H., & Gifford, J. A. (1986). Bayesian estimation in the three-parameter logistic model. *Psychometrika*. doi: 10.1007/BF02295598

Team, R. D. C., & R Development Core Team, R. (2016). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*. doi: 10.1007/978-3-540-74686-7

Teng, T.-l., & Taveras, M. (2004). COMBINING LIVE VIDEO AND AUDIO BROADCASTING, SYNCHRONOUS CHAT, AND ASYNCHRONOUS OPEN FORUM DISCUSSIONS IN DISTANCE EDUCATION. *J. EDUCATIONAL TECHNOLOGY SYSTEMS*. doi: 10.2190/XNPJ-5MQ6-WETU-D18D

Traphagan, T., Kucsera, J. V., & Kishi, K. (2010). Impact of class lecture webcasting on attendance and learning. *Educational Technology Research and Development*. doi: 10.1007/s11423-009-9128-7

Vieira, I., Lopes, A. P., & Soares, F. (2014). The Potential Benefits of Using Videos in Higher Education. In *Edulearn14 cd proceedings.*

Wise, S. L. (1991). The Utility of a Modified One-Parameter IRT Model With Small Samples. *Applied Measurement in Education*. doi: 10.1207/s15324818ame0402_4

Table 2
Empricial Study Model Result

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%	Exp(median)
θ_1	1.01	201	-0.60	0.06	-0.72	-0.60	-0.48	0.55
θ_2	1.02	129	0.71	0.05	0.62	0.71	0.81	2.04
θ_3	1.01	182	-0.13	0.06	-0.24	-0.13	-0.01	0.88
θ_4	1.01	164	0.13	0.06	0.03	0.13	0.24	1.14
θ_5	1.01	194	-0.23	0.06	-0.35	-0.23	-0.11	0.80
θ_6	1.01	193	0.09	0.06	-0.03	0.09	0.21	1.09
θ_7	1.01	214	0.11	0.06	-0.01	0.11	0.23	1.12
θ_8	1.01	165	-0.17	0.06	-0.28	-0.17	-0.06	0.84
θ_9	1.01	147	0.24	0.05	0.14	0.24	0.34	1.27
θ_{10}	1.01	162	0.01	0.05	-0.10	0.00	0.11	1.00
θ_{11}	1.01	172	-0.23	0.06	-0.34	-0.23	-0.12	0.80
θ_{12}	1.01	238	-0.47	0.07	-0.59	-0.47	-0.34	0.63
θ_{13}	1.01	156	0.03	0.05	-0.07	0.03	0.14	1.03
θ_{14}	1.01	162	0.14	0.06	0.03	0.14	0.25	1.15
θ_{15}	1.02	135	0.65	0.05	0.55	0.65	0.75	1.91
θ_{16}	1.01	161	-0.02	0.05	-0.12	-0.02	0.09	0.98
θ_{17}	1.01	155	0.02	0.05	-0.08	0.02	0.13	1.02
θ_{18}	1.01	148	0.17	0.05	0.07	0.17	0.28	1.19
θ_{19}	1.01	149	0.11	0.05	0.00	0.11	0.21	1.11
θ_{20}	1.01	162	-0.17	0.06	-0.28	-0.17	-0.06	0.84
θ_{21}	1.01	176	0.00	0.06	-0.11	0.00	0.12	1.00
θ_{22}	1.01	149	0.25	0.05	0.15	0.25	0.36	1.28
θ_{23}	1.01	234	-0.29	0.07	-0.41	-0.29	-0.16	0.75
θ_{24}	1.01	173	-0.31	0.06	-0.42	-0.31	-0.20	0.74
θ_{25}	1.02	134	0.52	0.05	0.42	0.52	0.61	1.67
θ_{26}	1.01	187	-0.20	0.06	-0.31	-0.20	-0.08	0.82
θ_{27}	1.01	162	0.03	0.05	-0.08	0.03	0.13	1.03
θ_{28}	1.01	159	-0.07	0.05	-0.17	-0.07	0.04	0.93
β_1	1.02	109	4.21	0.04	4.12	4.21	4.29	67.35
β_2	1.02	110	4.09	0.05	4.00	4.09	4.17	59.58
β_3	1.02	110	4.01	0.05	3.93	4.01	4.10	55.36
β_4	1.02	109	4.20	0.04	4.11	4.20	4.29	66.81
β_5	1.02	109	4.20	0.04	4.11	4.20	4.28	66.62
β_6	1.02	108	4.42	0.04	4.34	4.43	4.51	83.59
β_7	1.02	108	4.24	0.04	4.15	4.24	4.33	69.56
β_8	1.02	110	3.98	0.05	3.89	3.98	4.06	53.56
β_9	1.02	110	3.93	0.05	3.85	3.93	4.02	51.15
β_{10}	1.02	111	3.89	0.05	3.80	3.89	3.98	48.96
β_{11}	1.02	112	3.85	0.05	3.76	3.85	3.94	47.11
β_{12}	1.02	110	3.73	0.05	3.64	3.73	3.82	41.72
β_{13}	1.02	110	4.00	0.05	3.91	4.00	4.08	54.47

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%	Exp(median)
β_{14}	1.02	111	4.00	0.05	3.91	4.00	4.09	54.83
β_{15}	1.02	127	2.74	0.05	2.65	2.74	2.83	15.51
β_{16}	1.02	120	3.11	0.05	3.02	3.11	3.20	22.50
β_{17}	1.01	137	2.62	0.05	2.52	2.62	2.71	13.70
β_{18}	1.02	115	3.20	0.05	3.11	3.20	3.29	24.57
β_{19}	1.02	120	3.10	0.05	3.01	3.10	3.19	22.29
β_{20}	1.02	120	2.92	0.05	2.82	2.92	3.01	18.53
β_{21}	1.02	120	3.02	0.05	2.93	3.02	3.11	20.58
β_{22}	1.02	119	3.16	0.05	3.07	3.16	3.25	23.55
β_{23}	1.02	120	2.96	0.05	2.86	2.96	3.05	19.24
β_{24}	1.02	122	2.93	0.05	2.83	2.93	3.02	18.71
β_{25}	1.02	118	3.10	0.05	3.01	3.10	3.19	22.28
β_{26}	1.02	120	2.96	0.05	2.87	2.96	3.05	19.32
β_{27}	1.02	121	3.00	0.05	2.91	3.00	3.09	20.14
β_{28}	1.02	123	2.96	0.05	2.86	2.96	3.05	19.24
ϕ	1.00	21590	0.24	0.00	0.23	0.24	0.25	No need