Examining the Effect of Presenter Speaking Rate on Student Engagement with Educational Videos

Olivia Kendall; Daniel Tudball Smith; Mohamed Tolba; Alexander Gregg; Tony Vo; Scott Wordley; Michael Crocco
Monash University,
Corresponding Author Email: *olivia.kendall1@monash.edu*

CONTEXT
With increasing cohort sizes and a shift towards flipped classrooms, pre-recorded video content is heavily relied upon for student learning in undergraduate engineering courses. As a result, educators are under pressure to create engaging video resources without compromising on the depth and quality of unit material. Studies have explored strategies to achieve this by reducing video length, employing talking head video formats, and adopting a more informal presenter style. A high rate of speech is known to increase the extraneous cognitive load of the listener, which may negatively affect information comprehension (Mo et al., 2022). In comparison, slower talking speeds allow more time for information processing. This application of cognitive load theory has not been extensively investigated in educational videos.

PURPOSE
The purpose of this study is to investigate how the speaking rate, measured in words per minute, in engineering video content impacts the level of student engagement. This investigation aims to identify an optimal range of delivery speed to create videos to be used in engineering education.

APPROACH
The viewing analytics of over 250 pre-recorded lecture videos from four undergraduate engineering units across two universities have been critically assessed to quantify how speech tempo impacts student engagement. A correlation study was undertaken between the average speech rate and viewer retention metrics (a proxy for student engagement). A Multiple Linear Regression analysis was performed to include effects from video length, content type, and presentation style. In addition, video metrics have been used to estimate the playback speed used by viewers for varying speaking rate.

OUTCOMES
The level of student engagement was found to depend upon the rate of presenter speech within videos, though with influences from a number of other contributing factors, including video length, video format, content type, and presenter style. It was found that viewers typically increase playback speed to rates of 180-280 words per minute, with faster playback speeds being more common on videos that are presented slowly. From these outcomes, a series of recommendations are provided that highlight an optimal range of delivery speeds for short, mid and long-form video content to aid student learning.

CONCLUSIONS
The findings from this study will inform future research to determine the key parameters that impact student engagement and interaction with video content in engineering education. This will be used to identify best practices when developing learning resources for undergraduate engineers.

KEYWORDS
Student Engagement, Lecture Videos, Video Analytics

Introduction

Advances in digital technology, coupled with a growing demand for flexible education, have resulted in a widespread shift toward flipped classrooms where pre-recorded video content plays a central role in student learning. In undergraduate engineering programs, video-based resources are routinely employed to deliver lectures, provide supplementary explanations, worked examples and facilitate training, demonstrations and laboratory instruction (Resendiz-Calderón et al., 2024). As video software becomes more accessible to both students and educators, video-based learning offers a cost-effective solution to accommodate rising cohort numbers and help overcome size limitations in physical teaching spaces. In addition, videos offer enhanced accessibility by catering to diverse learning needs through features such as audio-visual delivery, subtitles and the ability to revisit material as needed. This aligns with a broader push for more flexible, student-centred education, enabling students to access content remotely and exercise autonomy over the rate and pace of their learning. Consequently, engineering educators are increasingly challenged to produce comprehensive video resources that convey complexity and depth while also capturing student engagement. As video-based learning reduces direct interpersonal interactions compared to traditional classroom settings, student engagement becomes especially critical for maintaining attention and motivation to achieve learning outcomes.

A framework for best practice to promote student interaction with video resources has been developed based on recommendations and evidence from several research initiatives. Shorter video durations, generally not exceeding ten minutes, consistently achieve higher audience retention compared to traditional lecture-length videos (Guo et al., 2014; Truss et al., 2024). Video formats that show a presenter onscreen, such as talking head styles, are more engaging for viewers in comparison to narrated slides (Dart, 2020). A high video production quality does not result in increased student engagement, as simply produced videos can be perceived as more personal and thereby relatable to viewers (Smith & Francis, 2022). In addition, the presenter's delivery style can positively influence viewer engagement as students show a preference for informal and enthusiastic delivery styles (Choe et al., 2019; Dart & Gregg, 2021).

Another critical aspect when developing video resources is the cognitive load placed upon the viewer. According to Cognitive Load Theory, it is the combination of intrinsic, extraneous and germane cognitive loads that impact learning (Sweller et al., 2011). Intrinsic load refers to the inherent difficulty of the learning material. Extraneous load is associated with how the material is presented, and germane load reflects the mental effort required to integrate the new knowledge into long-term memory. One of the main contributors to extraneous cognitive load in educational videos is the speaking rate of the presenter. Information that is presented too quickly cannot be processed and retained by the listener. Conversely, slow speaking rates increase the amount of information to be retained by the listener between instructions (Truss et al., 2024). In both cases, the working memory of the listener is overloaded, which prevents effective learning. Despite its importance, speaking rate has not been widely analysed in relation to student engagement with educational video content.

This study aims to begin addressing this gap by conducting a pilot investigation to examine the relationship between speaking rate and student engagement in higher education engineering videos. Relationships between delivery speed and viewer engagement will be used to inform the design of educational video resources for undergraduate engineers that better manage cognitive load and support student learning.

Purpose

This investigation explores the correlation between the presenter's speaking rate to student engagement in engineering videos using quantitative methods. For this investigation, student engagement with video content refers to increased viewing duration, and viewer retention will be used as a proxy for assessing the student engagement. Retention herein refers to the percentage of viewers watching the video at any given point in time. Therefore a higher viewer retention suggests greater engagement with the video content. Presenter speaking rate and audience retention will be examined in relation to other video characteristics such as video length, video type, and inferred viewer playback speed. The outcomes will assist in forming a speaking rate recommendation within the framework of best practice for designing effective educational videos for undergraduate engineers.

Methodology

Videos from four undergraduate Mechanical Engineering units across two Australian Universities were selected for this study and span first and second year undergraduate levels. Selection for videos from these units was based on the following criteria:

- Videos have been included in at least two separate offerings of the unit.
- The videos are hosted on YouTube with analytics available from the channel owners.
- The videos have been used in their respective units within the last 2 years.
- Each video presenter is included in at least 5 videos within the data set.

Based on these criteria, a total of 250 videos were analysed in this study using viewing metrics accessible from the YouTube hosting platform. A summary of video metrics for the videos analysed from each unit is shown below in Table 1.

Table 1: Summary of metrics from YouTube videos used in analysis, divided by speaker (S) and video type (WEx - worked example, Lec - lecture).

Identifier	Speaker	Unit	Number of videos	Total Views	Av. View Duration (mins)	Av. Video Length (mins)
S1-WEx	1	A	6	891	4.1	16.6
S2-WEx	2	B,C	39	54,830	6.9	14.1
S2-Lec	2	B,C	144	130,884	3.4	5.0
S3-Lec	3	D	22	208,573	3.4	10.9
S4-Lec	4	D	15	186,325	2.9	8.0
S5-WEx	5	D	24	17,576	2.0	6.3

Data Analysis

A Python script was used to analyse the video transcripts and determine the average speaking rate in words per minute (WPM) by dividing the total word count by video length minus video silent time. Each video was categorised according to speaker and video type ie. lecture or worked example, and video duration. An automated script, along with the YouTube data and analytics APIs was used to retrieve video metrics from the YouTube channel associated with each unit. The collected metrics include video duration, views, audience retention and average view duration. The output provided for audience retention is an array of 100 points, evenly distributed across the video timeline. Each value represents the percentage of the viewers watching the video at that point in the timeline and can reach above 100% if viewers rewatch specific sections of the video. A similar approach using audience retention data has previously been employed by various authors (Dart, 2020; Wordley et al., 2024; Guo et al., 2014). These studies also outline limitations in this approach, for example, students could play a video without paying attention or pause a video to work independently (Dart, 2020), and it is likely that other confounding factors affect how motivated a student is to watch or rewatch content. Therefore, future assessment of true student engagement requires a qualitative analysis of the student perspective. However for this preliminary quantitative investigation, audience retention is a useful proxy for assessing viewer engagement.

Multiple Linear Regression

Statistical analysis was conducted using multiple linear regression to investigate how speaking rate influences viewer retention within the dataset, along with other video characteristics. Average retention was taken as the dependent variable, whilst the independent variables included speaking rate measured in WPM, video duration, video style (lecture or worked example), video format (PowerPoint slides, Khan-style or talking head) and speaker. Interaction terms were also included to determine whether the effect of speaking rate varied depending on video length and video format. All continuous variables were standardised to assist with the comparison of effect sizes, and the statistical significance was evaluated at the $p < 0.05$ level.

Inferred Playback Speed

The playback speed multiplier (PBM) used by viewers is not available via the YouTube analytics API, however, the average playback speed can be inferred using other metrics. To determine the playback speed, it is required to first know the real-world watch time ('YouTube watch time') and the watch time relative to the video's progress bar ('video watch time'). The 'YouTube watch time' can be obtained directly from YouTube analytics either as a total time per video (t_Y) or as an average watch time per viewer on a given video (t_{Yav}). This is the time viewers spend watching the video, regardless of playback speed – e.g. 2 minutes of video watched at 2× speed will be recorded as 1 minute of 'YouTube watch time'. 'Video watch time' is not directly available from YouTube analytics, however, it can be inferred from the retention data. The retention, r, provided by YouTube analytics is given across the video's duration, is the proportion of viewers who are watching the video at any point. The number of viewers at any point in a video's progress, n_i, can then be obtained by multiplying the retention at that point by the total number of views, n_p. It then follows that the total 'video watch time' for all viewers (t_V) relative to the video's progress bar is the integral of n_i across the length of the video as measured by the video's progress bar, t_b.

$$t_V = \int n_i \, dt_b = n_p \int r \, dt_b$$

The average 'video watch time' per viewer is then simply,

$$t_{V\,av} = \int r \, dt_b \,,$$

This is visualised in Figure 1. This video watch time (either the total or the average) is the time that takes into consideration the playback speed – e.g. 2 minutes watched at 2x speed will be recorded as 2 minutes of watching. The inferred average playback speed multiplier is then the ratio between the time measured in these two reference frames.

$$PBM = \frac{t_{V\,av}}{t_{Y\,av}}$$

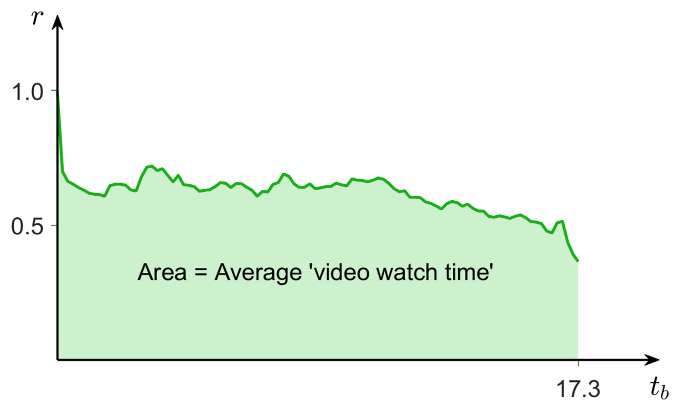

Figure 1. Typical retention vs video time plot for a 17.3 minute video.

Results and Discussion

A preliminary analysis of the dataset was conducted by categorising the videos into duration bins and plotting the average retention against average speaking rate, as shown in Figure 2. The graph shows a positive trend of average retention increasing with average speaking rate. Additionally, shorter video lengths, specifically those under six minutes, are clustered toward the higher end of average retention and speaking rate. This suggests that across the data set, shorter video content generally has a faster delivery rate, regardless of the presenter. Shorter videos also have higher viewer retention and student engagement compared to longer video formats. In contrast, longer videos, especially those over 10 minutes in duration, tend to be delivered with a slower average speaking rate with a lower average retention. This observed trend aligns with existing literature where videos below 10 minutes in length are more effective at maintaining audience attention (Guo et al., 2014; Truss et al., 2024). To better understand the effect of speaking rate on audience retention and video duration, this relationship was further examined through multiple linear regression analysis.

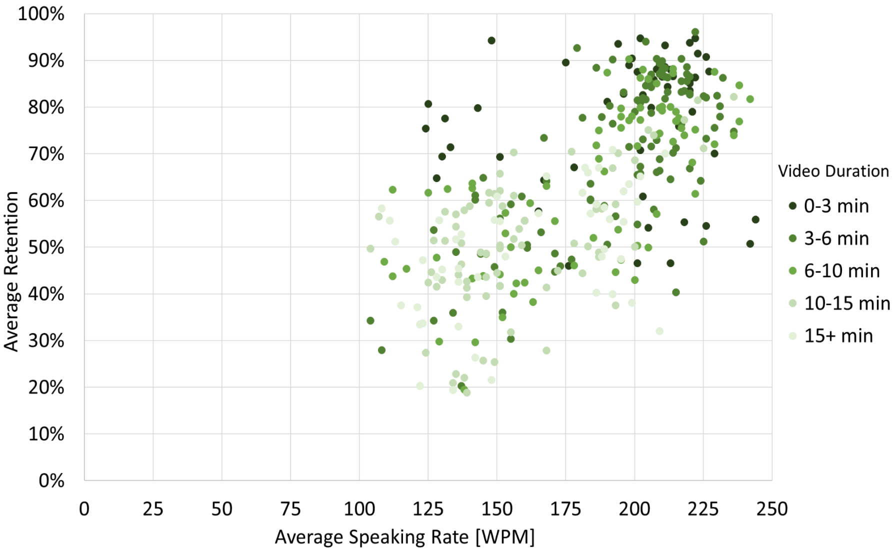

Figure 2. Average video retention plotted against average speaking rate, grouped by video duration.

Multiple Linear Regression Analysis

The results from the multiple linear regression analysis are summarised in Table 2 below. It can be seen that 71% of the variation in average retention can be explained by the independent variables, making the model statistically significant ($p < 0.001$). Using average values for speaking rate and video duration, a video-based lecture presented in a slide-based format is associated with an average viewer retention of 48%.

Table 2. Summary of results from multiple linear regression analysis

R Squared	0.71		
Variable	**Standardised Coefficient**	**p-value**	**t value**
Retention constant	0.48	0.00	21.01
Speaking rate^2	-0.11	0.00	-2.51
log (Video Duration)	-0.13	0.00	-5.47
log (Video Duration) x Speaking Rate	0.14	0.00	3.50
Worked Example	-0.10	0.00	-6.48
Worked Example x Speaking Rate	0.01	0.48	0.71
Khan-style	0.13	0.00	4.75
Talking Head	0.30	0.00	12.26
Fast Speaker	-0.03	0.29	-1.07

To better capture the non-linear effects of video duration, a log-transformed duration term was included. This duration predictor is statistically significant, indicating viewer retention is reduced with increasing video length, although at a diminishing rate. Speaking rate was also found to have a non-linear relationship with retention and was included as a squared term in the regression model. The inclusion of a squared speaking rate significantly improved the model fit and suggests retention improves with increased speaking rate up to an optimal point before declining. This is potentially due to cognitive overload due to reduced comprehension at very fast talking speeds. A significant interaction was found between speaking rate and log-transformed duration, with a coefficient of 0.14 and $p < 0.001$. A moderate increase in delivery speed can increase retention, even in longer videos, which may be due to increased cognitive engagement. These findings align with cognitive load theory as moderate presenter delivery speed, eliminating unnecessary pauses and gaps, may improve the pace of video content and reduce the extraneous load due to increased processing demands on viewers.

A worked example video format, which produces a coefficient of -0.10 and $p < 0.001$, slightly reduces viewer retention when compared to lecture content. This may be explained by students skipping sections in worked example videos to focus on specific steps or solutions. Talking head videos have the strongest positive effect on retention (coefficient of 0.30 and $p < 0.001$), which aligns with previous findings from Guo et al. (2014), where an on-screen presence significantly increased student engagement. Khan-style videos also have a moderate effect on increasing viewer retention, again due to increased interaction with the presenter. Using an above-average speaking rate to deliver video content shows a small negative trend, however is not significant in this model, given other factors which may affect speaking rate, such as content difficulty, video length and presenter delivery traits (ie. tone).

The finding that higher speaking rates are associated with increased viewer retention aligns with outcomes from Guo et al.(2014). Similarly, Shoufan (2019) identified that educational video popularity increased with higher presenter talking rates, making voice and speaking rate significant factors for a student's liking or disliking a video. However, a faster speaking rate alone is not sufficient to significantly influence viewer retention. It acts in combination with investigated parameters such as video duration, video style – particularly talking head formats and lecture or worked example formats. In addition, it also varies with other parameters harder to quantify, such as presenter enthusiasm, vocal tone and content complexity, which influences cognitive engagement and viewer attention. This agrees with findings from Díaz & Recabarren (2025), who identified that students have distinct preference groups regarding speaking rates in educational videos, and there is no universal presenter speaking rate that increases student attention. From these outcomes, the parameters of speaking rate, video duration, lecture and worked example formats, as well as speaker identity, were selected for a more detailed analysis of preferred audience playback speed.

Inferred Playback Speed Analysis

After identifying the main parameters that influence student engagement or audience retention using multiple linear regression analysis, further investigation has been carried out to relate these parameters to the preferred video playback speed. In this section, the average playback speed of a video is compared to the duration and words per minute of videos for different speakers, and whether that speaker is presenting lecture content or a worked example. The inferred average playback speed plotted against both duration and words per minute is shown in Figure 3 (left). A general trend emerges of viewers favouring an increased playback speed for videos with longer duration and lower words per minute. A viewer is likely to increase the speed of a low-speaking-rate video, but if that video is also short, on average it will be less likely to be sped up by viewers. This trend applies when considering all speakers, units, and delivery types together, but the trends are less obvious when considering those combinations individually. The average playback speed is also plotted against the effective duration and words per minute in Figure 3 (right). These ‘effective’ values are obtained after taking into account the average playback speed (e.g., to a viewer watching at 2x speed, the video duration is halved and words per minute is doubled). It is observed from this figure that the effect of applying the playback speed is to pull the data closer together in terms of shortening the duration, and shift up the average words per minute.

This is further highlighted in the histograms presented in Figure 4. Presenters in the dataset mostly deliver content between 130 and 230 WPM (average of 195, standard deviation of 29), but on average, the videos are being viewed at 180 to 280 WPM (average of 240, standard deviation of 29). The spread of speaking rates remains similar, as although some slower videos have higher playback speeds bringing them closer to the average, other videos are sped up to take them beyond the average, resulting in an almost identical standard deviation before and after accounting for the average playback speed. For video duration, since playback speed is mostly in the 1.1 to 1.5 range, the histogram is compressed towards shorter durations.

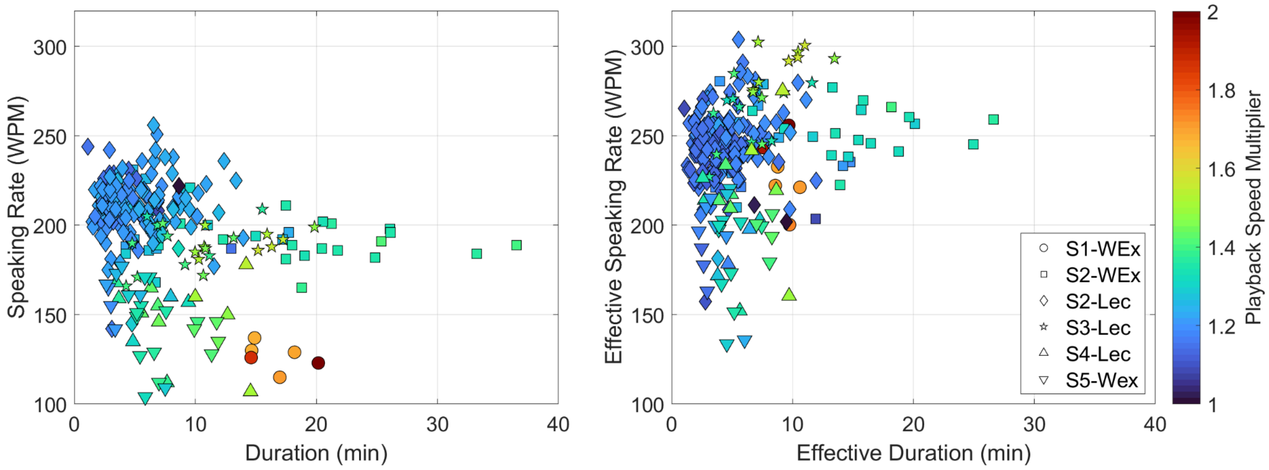

Figure 3. The inferred average playback speed plotted against words per minute and video duration for (left) the raw WPM and duration and (right) the effective WPM and Duration after multiplication and division, respectively, by the inferred average playback speed.

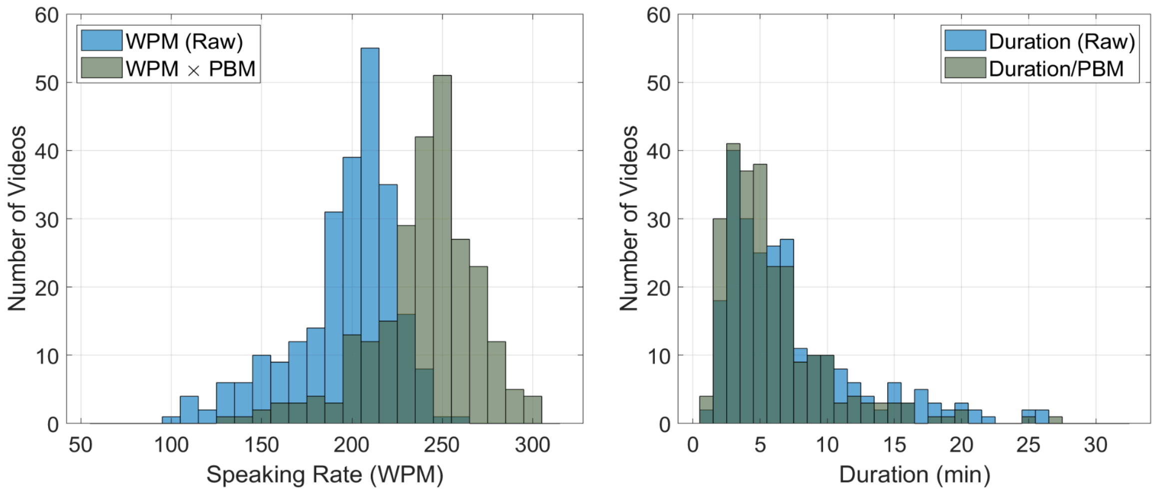

Figure 4. Histograms of the (left) words per minute and (right) duration of a video before ('raw') and after multiplication and division, respectively, by the average playback speed multiplier.

Previous trends apply when considering all speakers, units, and delivery types together, but individual trends are less clear. Figure 5 presents the playback speed against words per minute (left) and video duration (right) for each presenter-type combination. Considering words per minute (Figure 5 left), a general trend of increasing playback speed with decreasing words per minute can be observed. For Speaker 2 presenting lecture content, over a broad range of WPM there is minimal variation in playback speed, indicating viewers found a comfortable speed and maintained it throughout the content. However, for the same speaker presenting a worked example, the playback speed becomes greater with reduced speaking rate. Speakers 1, 3 and 5 also show trends of increasing playback speed with decreasing speaking rate for both lectures and worked examples. This increase varies between speakers as indicated by clusters in different areas of the plot. All speaker-format combinations have a positive trend with duration (Figure 5 right), but Speaker 5, for example, has a larger gradient than Speaker 2 for worked example videos - indicating audiences prefer higher speaking rates for Speaker 5 than Speaker 2 for the same video length. Speaker 3 has a similar gradient to Speaker 2 for lecture content, but the y-intercept is greater – indicating a potential higher baseline playback speed being preferred for Speaker 3. So while it can generally be seen that playback speed is increased for increasing duration and decreasing talking speed, the specific magnitude of the increased playback speed will be dependent on other factors. Other contributing factors may include presenter enthusiasm, content complexity, as well as a perceived increase in productivity associated with watching videos at higher speeds, although it remains unclear where the balance lies between comprehension and productivity. Viewers may aim to consume a large amount of content in a short time, even if it is not well understood.

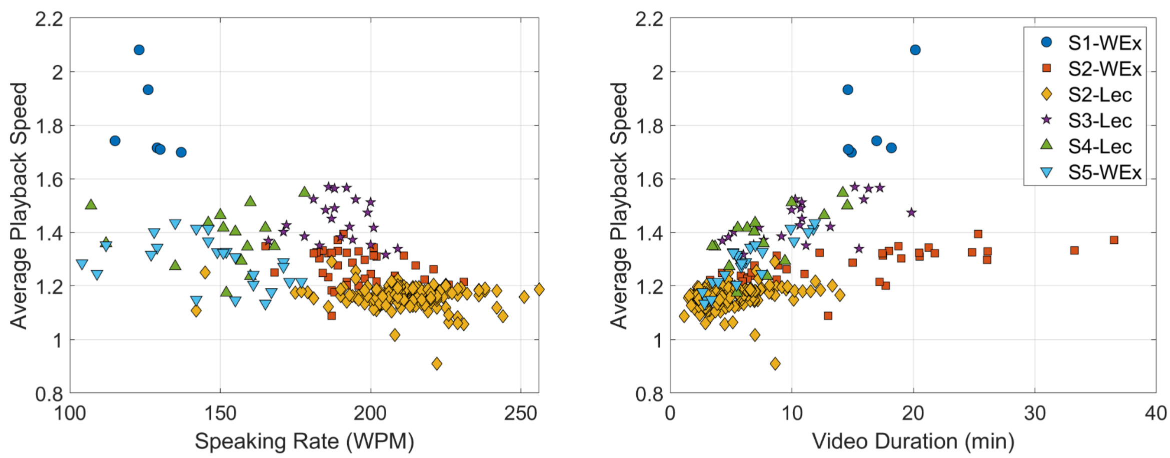

Figure 5. The average inferred playback speed plotted against (left) the words per minute and (right) the video duration for a range of videos from different speakers and content types.

Limitations

This pilot study analysed 250 videos from four engineering undergraduate units. This sample contains two primary video styles, two video formats and five different presenters, with video durations ranging from 32 seconds to 43 minutes. Future study and analysis are required to expand the dataset to include a broader range of units, speakers and delivery formats to ensure outcomes are more generalised. In addition, a qualitative investigation will be carried out to complement the quantitative analysis. This will examine student perceptions of video content, video engagement and information retention and will further explore the relationship between speaking rate and viewer cognitive load in engineering videos. Some caution should be taken when interpreting this data. Overall trends seen across the entire dataset occasionally vary compared to those for individual speakers or formats. Though generally the trend itself is similar, the magnitudes of those trends may vary. Individual speakers tend to have a preferred style (e.g. talking speed and may favour longer/shorter videos). This results in individual combinations tending to be clustered within the overall dataset and makes it difficult to ascertain whether findings are globally valid. So while this work has presented some broad trends which can form recommendations, the individual effect on a video is likely to be influenced by other factors as discussed above.

Conclusions and Recommendations

The outcomes from multiple linear regression and inferred playback speed analysis indicate that speaking rate does influence student engagement with educational video resources. The main findings from this study are:

1. A faster speaking rate increases viewer retention to an optimal point, after which retention declines. When filming longer videos, presenters should consider adopting a slightly faster speaking rate to promote student engagement.

2. Almost all videos have a playback speed greater than 1.1, indicating students are likely to increase video speeds between 180 to 280 words per minute on average. It is not necessarily recommended for presenters to attempt to match these speaking rates, as student behaviour shows that playback speed will still be increased, even for fast-paced presenters.

3. Factors such as personal delivery style, tone and enthusiasm, as well as course difficulty, are as significant as speaking rate and video duration in affecting student engagement. Presenters are recommended to use a speed and delivery style that is comfortable to them and may consider using talking head video formats when producing lectures and worked examples.

References

Choe, R. C., Scuric, Z., Eshkol, E., Cruser, S., Arndt, A., Cox, R., Toma, S. P., Shapiro, C., Levis-Fitzgerald, M., Barnes, G., & Crosbie, R. H. (2019). Student Satisfaction and Learning Outcomes in Asynchronous Online Lecture Videos. *CBE Life Sciences Education*, *18*(4), ar55–ar55. https://doi.org/10.1187/cbe.18-08-0171

Dart, S. (2020, January 1). *Khan-style video engagement in undergraduate engineering: Influence of video duration, content type and course* (informit.723741887771795). 31st Annual Conference of the Australasian Association for Engineering Education (AAEE 2020): Disrupting Business as Usual in Engineering Education, Barton, ACT. https://search.informit.org/doi/10.3316/informit.723741887771795

Dart, S., & Gregg, A. (2021, January 1). *Know your stuff, show enthusiasm, keep it on message: Factors influencing video engagement in two mechanical engineering courses* (informit.345896810873543). REES AAEE 2021 conference: Engineering Education Research Capability Development, Perth, WA. https://search.informit.org/doi/10.3316/informit.345896810873543

Díaz, M., & Recabarren, M. (2025). *Beyond one-size-fits-all: Speech rate personalization as a key to inclusive video lecture design*. *Universal Access in the Information Society, 24*(1), 747–758.https://doi.org/10.1007/s10209-024-01106-1https://doi.org/10.1007/s10209-024-01106-1

Guo, P. J., Kim, J., & Rubin, R. (2014). How video production affects student engagement: An empirical study of MOOC videos. *Proceedings of the First ACM Conference on Learning @ Scale Conference*, 41–50. https://doi.org/10.1145/2556325.2566239

Mo, C.-Y., Wang, C., Dai, J., & Jin, P. (2022). Video Playback Speed Influence on Learning Effect From the Perspective of Personalized Adaptive Learning: A Study Based on Cognitive Load Theory. *Frontiers in Psychology*, *13*, 839982–839982. https://doi.org/10.3389/fpsyg.2022.839982

Resendiz-Calderón, C. D., Farfan-Cabrera, L. I., Cazares-Ramírez, I. O., Nájera-García, P., & Okoye, K. (2024). Assessing benefits of computer-based video training and tools on learning outcomes and motivation in mechanical engineering education: Digitalized intervention and approach. *Frontiers in Education*, *Volume 9-2024*. https://www.frontiersin.org/journals/education/articles/10.3389/feduc.2024.1292405

Shoufan, A. (2019). *What motivates university students to like or dislike an educational online video?*. Computers & Education, *134*, 132–144. https://doi.org/10.1016/j.compedu.2019.02.008

Smith, D. P., & Francis, N. J. (2022). Engagement with video content in the blended classroom. *Essays in Biochemistry*, *66*(1), 5–10. https://doi.org/10.1042/EBC20210055

Sweller, John., Ayres, Paul., & Kalyuga, Slava. (2011). *Cognitive Load Theory* (1st ed. 2011.). Springer New York. https://doi.org/10.1007/978-1-4419-8126-4

Truss, A., McBride, K., Porter, H., Anderson, V., Stilwell, G., Philippou, C., & Taggart, A. (2024). Learner engagement with instructor-generated video. *British Journal of Educational Technology*, *55*(5), 2192–2211. https://doi.org/10.1111/bjet.13450

Wordley, S., Gregg, A., Jones, L. J. N., & Vo, T. (2024, December). *Make shorter videos: Insights into educational video duration from quantitative analytics* (informit.T2025032400012901790470017). Proceedings of the 35th Annual Conference of the Australasian Association for Engineering Education (AAEE 2024), Christchurch, New Zealand. https://search.informit.org/doi/10.3316/informit.T2025032400012901790470017

Copyright statement