The Impact of Instructor Personality on the Success of Online Educational Videos: A Multimodal Deep Learning Approach

Yangfan Liang
CMU Heinz College, yangfanl@andrew.cmu.edu

Mi Zhou
UBC Sauder School of Business, mi.zhou@sauder.ubc.ca

Pedro Ferreira
CMU Heinz College, pedrof@cmu.edu

Michael D. Smith
CMU Heinz College, mds@cmu.edu

ABSTRACT

Online education is a vital consumer industry that is undergoing rapid technological change. Despite the growth of online education, student engagement and retention rates online have lagged relative to physical classrooms. Yet, the determinants of online educational video success remain mostly unexplored, due in part to the difficulty of analyzing unstructured video data. In this study, we analyzed a unique large-scale video dataset to investigate the impact of instructor personality on the success of online educational videos. Specifically, we first propose a novel multimodal deep learning model to measure an instructor's latent personality traits from video content (i.e., image, audio, text), finding that visual signals contained in images provide more relevant information for measurement of personality traits than do linguistic information in subtitles or acoustic signals in sound. We then empirically examine the impact of instructor personality on the performance of online educational videos, leveraging the double machine learning method to tease out potential confounding effects of an extensive set of video features. Our results show that the level of extraversion of an instructor has a positive and statistically significant effect on the course video's performance, whereas openness is negatively associated with video performance. We also find that the impact of latent personality traits is moderated by an instructor's observable characteristics such as gender and age. Our paper provides managerial implications for online education platforms with respect to their efforts to improve digital product design and enhance user engagement with online video content.

Key words: online education, instructor personality, multimodal deep learning, video analytics

1. Introduction

Education is a vitally important industry, both economically and socially — and one that is being transformed by the information technology. According to recent statistics, education spending represents 6.1% of Gross Domestic Product (GDP) in the United States and 5.0% of GDP worldwide (Investopedia 2019). In the context of the transformation of the education industry by technology, in 2019 the global online education market was estimated at almost $200 billion, and is predicted to reach $400 billion by 2026 (Statista 2022). Moreover, the COVID-19 pandemic, created arguably the largest disruption of education systems in history (UnitedNations 2020), causing a surge in the adoption of online education that is unlikely to recede. The rise of online education presents scholars with a unique opportunity to directly observe data on individual behavior and uncover novel insights into this economically and culturally important industry.

Despite the advancement and convenience brought by online education, relatively little is known about the characteristics that affect individual student behavior and outcomes when consuming educational materials online. Moreover, student engagement with online courses and retention is reported to be much lower than in the case of physical classrooms. For example, reports show that the average completion rate for Massive Open Online Courses (MOOCs) is as low as 15% (Jordan 2015). Unlike traditional education, where learning occurs in physical classrooms, online education usually relies on video courseware to deliver educational materials. Yet, the determinants of the success of online educational videos remain mostly unexplored, largely due to the unique challenges entailed in generating insights from unstructured video data.

More broadly, digital video consumption has been on the rise, as witnessed by the growing popularity of online video platforms such as YouTube, TikTok, and Netflix. By the end of 2022, online videos make up more than 82% of all consumer Internet traffic, which is a 15-times larger proportion than for 2017 (InVideo 2022). As many as 78% of people report watching videos online each week, and 55% of them watch online videos on a daily basis (Perry 2019). Importantly, people watch online videos not only for entertainment purposes, but also for education and learning. Recent

statistics show that learning and educational content drives over a billion views a day on YouTube (Wojcicki 2018), and corporate employees are reportedly 75% more likely to watch a video than to read documents or web articles (TechSmith 2022). Despite the growing popularity and importance of online video consumption in today's digital economy, there is a lack of research on video analytics in the information systems (IS) literature, which presents important new opportunities for researchers, particularly given the abundance of available video data and recent advancements in big data and deep learning techniques.

Our research aims to address these gaps in the online education literature and video analytics literature by examining, based on theories in psychology and social sciences, whether and how an instructor's latent personality (as extracted from video content) affect the success of online educational videos. Personality theories have been used to understand individual behavior in different contexts such as political science (Gerber et al. 2011), entrepreneurship (Antoncic et al. 2015), e-government portal use (Venkatesh et al. 2014), technology adoption (Devaraj et al. 2008), and consumer behavior (Liu et al. 2016). Simply put, automated personality detection can provide rich predictors, informing an array of downstream analytics applications (Yang et al. 2022). Although leveraging personality traits is a promising pathway toward understanding individual behaviors in various settings, there exist significant challenges that have so far prevented the full exploitation of personality characteristics as predictors of individual behaviors on a large scale, due to the paucity of available psychometric data (Ahmad et al. 2020) and the inherent difficulty of measuring latent personality characteristics (Adamopoulos et al. 2018). Specifically, the traditional way of measuring personality characteristics requires subjects to complete long questionnaires, making it particularly burdensome, if not impossible, to obtain such information on a large scale. More recently, a few studies in the IS literature have succeeded in automatically assessing personality traits from text data, and have shown that personality traits are associated with the effectiveness of word-of-mouth (Adamopoulos et al. 2018), streamers' popularity (Zhao et al. 2019), review helpfulness (Liu et al. 2021), and firms' financial outcomes (Yang et al. 2022).

Our paper extends this stream of research in the IS literature in two novel ways. First, we complement this literature by looking into an entirely different market: online education, a growing sector of increasing economic and societal importance. Examining the impact of instructors' latent personality traits on the performance of their online educational videos provides important managerial insights for both instructors and online education platforms, which will also benefit students and society at large. Second, whereas many studies have inferred personality traits based on text data, our study extends this literature methodologically by proposing a multimodal model to infer personality traits using video data based on textual, visual, and auditory information simultaneously.

This multimodal perspective is particularly important in our research context, which is to say, online education, wherein face-to-face human interactions are lacking. Importantly, when people watch videos online they are likely to be affected by linguistic (e.g., subtitles), visual (e.g., images), and acoustic (e.g., audio) messages simultaneously. Moreover, signals from each modality may not contribute equally to the video's perceived meaning, as these signals tend to be interconnected and interdependent. For example, visual messages in a video (e.g., facial expressions) or acoustic messages (e.g., tone, accent) might convey unique information about people's hidden characteristics that is hard to infer from caption text alone. Therefore, ignoring the impact of visual or acoustic signals, which represent distinct information on the video content, may lead to biased results. In our study, we address this issue by developing a multimodal personality model based on all three modalities (i.e., image, audio, text) for accurate inference of personality traits in video data.

Specifically, we first propose a novel multimodal deep learning model to automatically measure latent personality traits from video content (i.e., image, audio, text) and analyze the contribution of each modality to the prediction. We then empirically examine the impact of instructor personality on the performance of online educational videos using the double machine learning (DML) framework to tease out potential confounding effects of an extensive set of video features. To address our research questions, we collected a unique large-scale video data consisting of 10,000 videos

with labeled personality traits for model training and 13,869 online course videos from YouTube education channels for our primary analysis. These online course videos were gathered from six major education channels on YouTube, including Crash Course, SciShow, MIT OpenCourseWare, YaleCourses, Stanford Online, and UCI Open. These six channels have a total of more than 26 million subscribers on YouTube. In addition to the course videos and subtitles from YouTube, we also acquired, from a third-party company, a dataset that captures the historical views and likes of these videos on a daily basis. Using these datasets, we built a multimodal deep learning model to infer latent personality traits based on visual, linguistic, and acoustic information simultaneously. We found that our multimodal approach consistently outperformed unimodal models relying on a single modality (e.g., text). Moreover, our analysis reveales that visual messages in the video had larger weights for measurement of personality traits than for linguistic or acoustic messages. Additionally, we extracted a rich set of theory-driven features for each video as control variables, including basic video properties and visual aesthetic features as well as instructor's emotions and appearance features. Using the DML framework to account for the potential confounding effects of these variables, we found that the level of extraversion of an instructor is positively associated with video performance, whereas the level of openness of an instructor is negatively associated with performance. Furthermore, we found that these effects of latent personality traits are moderated by the instructor's observable characteristics such as age and gender.

These results make a novel contribution to the literature by combining deep-learning-based video-mining techniques with empirical analysis, and being the first to show how an instructor's latent personality traits affect the success of online educational videos as measured by consumers' likes and views. We believe these results and our proposed multimodal deep learning model pave the way for additional research in the underexplored area of online education and digital video consumption analytics. More specifically, by integrating and modeling multiple communicative modalities in video content, including linguistic, acoustic, and visual messages, our research provides actionable insights into the future potential of online education and business analytics using big data and unstructured video analysis.

The remainder of this paper is organized as follows. In Sections 2 and 3, we discuss the relevant literature and describe our data. In Section 4, we propose a multimodal deep learning model and evaluate its prediction performance. In Sections 5 and 6, we introduce our empirical analysis and discuss the corresponding results, and conclude.

2. Literature Review

In this section, we discuss how our study is related to, and extends, different streams of research in the online education market, drawing on the personality theories that are deeply rooted in psychology and the social sciences.

2.1. Online Education Market

The consumption and provision of education is one of the most resource intensive and consequential activities for providers and consumers worldwide (Stone 2018). Education is the latest industry to face digital disruption (Dellarocas and Van Alstyne 2013), and several technological innovations have enabled the emergence of online education. Driven by these technological innovations, the online education market is expected to grow rapidly, potentially reaching a value of $400 billion by 2026 (Statista 2022). Yet, despite its economic and social importance, online education has received relatively little attention in the IS or the management literatures (Zhang et al. 2017).

There are as yet only a small number of studies on this burgeoning literature. Adamopoulos (2013) investigates the determinants affecting student retention in online courses. Dellarocas and Van Alstyne (2013) focuses on business models for MOOCs. Terwiesch and Ulrich (2014) examines the emergence of MOOCs and their impact on business schools. Chen et al. (2016) investigates the relationship between student personality and learning behavior. Zhang et al. (2017) studies the impact of social interaction on students' online learning outcomes. Matcha et al. (2020) examines the relationship between learning strategy and personality traits. Huang et al. (2021) investigates the efficacy of alternative informational interventions for reducing users' procrastination in MOOCs. Leung et al. (2022) investigates the advantages of gamification for learners' engagement and learning outcomes in MOOCs.

Our research extends this literature by drawing on established theoretical concepts in psychology to examine the impact of instructors' latent personality traits of instructors on the success of online educational videos. Online educational videos have become a major format of education media, especially during the COVID-19 pandemic. Most governments around the world having temporarily closed educational institutions to contain the pandemic, nationwide closures have impacted over 60% of the world's student population (UNESCO 2022). Most institutions have facilitated the continuity of education through online learning, wherein video plays a critical role in the communication between instructors and students. Thus, our paper makes an important contribution to the literature by investigating the impact of an instructor's latent personality traits on the success of online educational videos in this emerging, economically and socially important customer service sector.

2.2. Personality Theories and Relevant Work

Personality encompasses a set of characteristics according to which a person thinks, feels, and behaves (AmericanPsychologicalAssociation 2022). Personality has been a widely studied field in psychology (Matthews et al. 2003), where several frameworks have been proposed to characterize personality—the Big Five personality model being the most influential and recognized (Costa et al. 1991, Goldberg 1990). This model proposes a comprehensive theoretical framework of five factors necessary and sufficient to represent human personality in terms of the traits that distinguish, order, and name the behavioral, emotional, and experiential characteristics of individuals (John et al. 1999). This model identifies personality across the following five dimensions: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (OCEAN). Openness measures one's tendency towards new experience; one with high openness, therefore, is more willing to try new things. Conscientiousness measures one's tendency towards self-discipline, determination, and achievement. Extraversion measures an individual's preferences toward stimulation from the outside world. Agreeableness measures one's tendency to get along with other people. Neuroticism measures one's tendency to have negative emotions. The Big Five personality model has been

extensively used to understand individual behavior in different fields including political science (Gerber et al. 2011), entrepreneurship (Antoncic et al. 2015), e-Government use (Venkatesh et al. 2014), technology adoption (Devaraj et al. 2008), and consumer behavior (Liu et al. 2016). Prior research has shown that personality traits are significant and powerful predictors of future outcomes that have policy implications (Yang et al. 2022).

In education, researchers have primarily focused on understanding student personality (Kim et al. 2019). For example, many studies have shown that students' personality traits are associated with their academic achievement (Richardson et al. 2012), academic motivation (Komarraju et al. 2009), academic honesty (Giluk and Postlethwaite 2015), and career decision-making (Martincin and Stead 2015). However, there is still a lack of studies on teacher personality, especially studies using established personality theories (Kim et al. 2019). Moreover, given the significant differences in many aspects between traditional education and online education, such as delivery mode, schedule, pace, flexibility, cost, and human interaction, the findings from the offline environment might not be generalizable to the online context.

More recently, personality traits have been studied in the IS literature to better understand consumer behavior in digital contexts. Specifically, Adamopoulos et al. (2018) used text mining to infer personality traits of users based on Twitter posts, and found that the level of personality similarity between social media users has a significant positive impact on the effectiveness of word of mouth messages. Zhao et al. (2019) focus on livestreaming content, using text mining to infer the personality traits of streamers based on their Twitter posts. They found that the personality traits of streamers were significantly correlated with their popularity on Twitch. Liu et al. (2021) apply text mining to Yelp reviews to derive the personality traits of reviewers, finding that their personality traits are associated with review helpfulness. Yang et al. (2022) develop a new method to extract personality traits from textual data, and show its effectiveness in predicting downstream tasks in the finance and health domains. We note that all of these studies rely on unimodal information (i.e., text) to infer people's personality traits. Recent work has considered this to be a

limitation, suggesting that it would be beneficial for future work on personality to use multimedia inputs including audio and video (Yang et al. 2022). This is one of the inspirations behind the present study.

3. Data

For our main analysis we collect a unique large-scale video dataset from two sources: 10,000 videos from the First Impression dataset (Ponce-López et al. 2016) with labeled personality traits for model training, and 13,869 online educational videos from YouTube.

To do this we first used the First Impression dataset to train our multimodal prediction model to automatically measure personality traits in each video in a scalable manner. The First Impression dataset, introduced in 2016, is the most popular multimodal dataset with labeled personality traits. It consists of 10,000 videos, specifically high-definition 15-second video clips from YouTube. It includes a 3:1:1 train/validation/test split, 6,000 videos having been assigned to the training set, 2,000 to the validation set, and 2,000 to the testing set. The 15-second video clips are question and answer videos where people talk to the camera with a clear voice and single face occurrence. Each video clip is paired with its caption and ground truth labels for the five personality traits, where the labels were derived from the pair-comparison results for Amazon Mechanical Turkers using a Terry-Luce model (Ponce-López et al. 2016, Bradley and Terry 1952). Because "Neuroticism" is the only negative trait, Ponce-López et al. (2016) replaced it with its opposite (non-Neuroticism, which is also called "emotional stability" in the literature) to score all traits in a similar way on an positive scale in the First Impression dataset. We follow this practice in our analysis. Each of these five personality traits is a continuous variable ranging from 0 to 1.

Second, our primary dataset consists of 13,869 online course videos from six major educational channels on YouTube, including Crash Course, SciShow, MIT OpenCourseWare, YaleCourses, Stanford Online and UCI Open. In total, these six channels have more than 26 million subscribers and offer a wide range of courses including statistics, computer science, physics, history, chemistry, biology, and ecology. We also obtained a time-coded subtitle file for each video in our data. In

addition to the course videos, we acquired, from a third-party company, a dataset that captures the historical views and likes of these videos. With these datasets, we identified the personality traits of each instructor in the video using our proposed multimodal deep learning model, which is introduced in the next section. Further, we extracted a rich set of theory-driven video features for each video in our data in order to control for potential confounding effects. We then used the DML framework to estimate the impact of instructor personality on the performance of the video based on the number of likes and views in the historical records.

To ensure that we could accurately detect instructors' personality traits in the video, we conducted a filtering process for data cleaning purposes. Specifically, we removed videos lacking historical views or likes data along with videos lacking the instructor's face. This data cleaning procedure left us with 6,090 videos. When measuring the personality traits of the instructor in the video, we extracted, for each video, a 15-second clip wherein the instructor's face appears as well as subtitles corresponding to the content of the clip. In Figure 1, we present several sample video frames showing different instructors.

Figure 1 Screenshots of Different Instructors with Different Personalities

4. A Multimodal Deep Learning Model to Measure Personality

Our empirical strategy consists of two components: first, automatic and scalable identification of the instructor's personality traits in each video using our proposed multimodal deep learning model; second, detection of personality traits combined with econometric analysis to investigate whether and how instructors' personality traits affect the success of online educational videos.

To automatically identify the personality traits of instructors in video data, we propose a multimodal deep learning model for prediction of each of the five personality traits based on the visual, auditory, and textual signals simultaneously. This approach has several advantages. First, machine-learning-based automated methods for personality assessment are more efficient and objective than traditional ways of measuring personality (Adamopoulos et al. 2018). Second, the traditional way of measuring personality, which requires people to complete personality questionnaires, does not allow for obtaining personality traits on a large scale or at low cost for the population of interest (Chen et al. 2015). Third, user-generated content is more reflective of users' actual personalities than their own "self-idealization" (Back et al. 2010).

In this section, we first introduce three unimodal models that use only one source of information as input (i.e., either image, audio, or text) to select the best candidates. Then, we introduce our multimodal deep learning model, which fuses all three information-signal sources to yield the best prediction performance. In our study, after we trained the model on the First Impression dataset, we analyzed the contribution of each modality to the final predictions in order to assess the impact of each (i.e., of text, image, and audio) on the personality trait prediction.

Since our objective variables were continuous, our models were evaluated by the mean accuracy of each video's personality prediction. The mean accuracy for the j-th personality trait was defined as $A_j = 1 - \frac{1}{N_t}\sum_{i=1}^{N_t}|t_i - p_i|$, and the mean accuracy over the five personality traits was defined as $A = \frac{1}{5}\sum_{j=1}^{5} A_j$, where N_t denotes the number of samples in the test set, t_i the true value, and p_i the predicted value. Next, we introduce each of the three unimodal models and our multimodal model and then compare their performances.

4.1. Text-Based Model

We first built our text-based model based on BERT (Bidirectional Encoder Representations from Transformers) so as to predict the personality traits using the subtitles of each video as input. The model is schematized in Figure 2. BERT is a method developed by Google to pretrain language models to solve a wide range of downstream natural language processing tasks (Devlin et al. 2018). First released in 2018, it initially produced state-of-the-art results on 11 different natural language processing tasks. Following BERT, several similar methods have been proposed, among which are XLNet (Yang et al. 2019) and RoBERTa (Liu et al. 2019). We chose BERT for our setting because of its high performance and advantages. First, the transform module it uses is capable of capturing the contextual information of each word. Specifically, given different contexts, it can capture different meanings of the same word. Second, it is pretrained on two specific unsupervised tasks, Masked Language Model (Masked LM) and Next Sentence Prediction (NSP), which enable it to capture information from both directions. More details on BERT are included in Appendix A. For our baseline used for comparison, we chose TF-IDF on n-grams in order to vectorize our text data into suitable features, and then used Random Forest to make predictions.

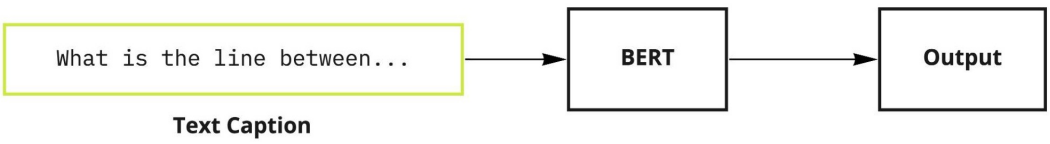

Figure 2 Schematization of Text-Based Deep Learning Model

Table 1 summarizes the performances of the text-based models. As expected, all three more advanced models (BERT, XLNet, and RoBERTa) outperformed the baseline model, which uses n-grams and Random Forest. The differences in the performances of those three models were negligible (all around 0.89). We thus selected BERT as the candidate of text-based model for the joint multimodal model.

	n-grams + Random Forest	BERT	XLNet	RoBERTa
Mean Accuracy	0.8737	0.8899	0.8869	0.8899
Extraversion	0.8725	0.8858	0.8817	0.8859
Non-Neuroticism	0.8657	0.8866	0.8840	0.8864
Agreeableness	0.8864	0.8989	0.8967	0.8989
Conscientiousness	0.8671	0.8865	0.8836	0.8871
Openness	0.8765	0.8915	0.8883	0.8914

Table 1 Performances of Text-Based Models

4.2. Image-Based Model

For the purposes of our image-based models, we built our deep learning models by fine-tuning EfficientNet B0 (Tan and Le 2019). We also used MobileNet V2 (Sandler et al. 2018) as a baseline along with a larger variant of EfficientNet, EfficientNet B7, for comparison. EfficientNet is a series of convolutional neural network structures that has been used for by its unique scaling utilities. There are three major factors in scaling convoluted neural networks: width, depth, and resolution. Traditionally, scaling methods tend to optimize one of those factors. However, it is intuitive that the factors need to be optimized jointly, since a higher-resolution model requires a deeper network to analyze. Thus, EfficientNet uses a compound coefficient to scale width, depth, and resolution jointly, and has achieved better accuracy and efficiency compared with previous models. EfficientNet B0 is derived from MobileNet V2 and is the smallest network in the EfficientNet family. It requires 0.39 billion FLOPS to train and achieves 77.1% top-1 accuracy on ImageNet. By comparison, MobileNet V2 requires 0.3 billion FLOPS to train but only achieves 72.0% top-1 accuracy. EfficientNet B7 is the largest model in its family, needing 37 billion FLOPS to train, achieving 84.3% top-1 accuracy. We selected EfficientNet B0 as the basic module for image processing, due to its balance between efficiency and performance. For each video, we extracted one video frame per second. Since we analyzed 15-second video clips, this generated 15 frames in total for each video. For each video frame, we first passed it through a pretrained model (e.g., EfficientNet B7, EfficientNet B0, or MobileNet V2), and then passed the outputs into a long short-term memory (LSTM) model to

capture potential time dependencies. Finally, the output from the LSTM model was passed through two fully connected layers to derive the predictions of each of the five personality traits. More details about EfficientNet are shown in Appendix B. This process is schematized in Figure 3.

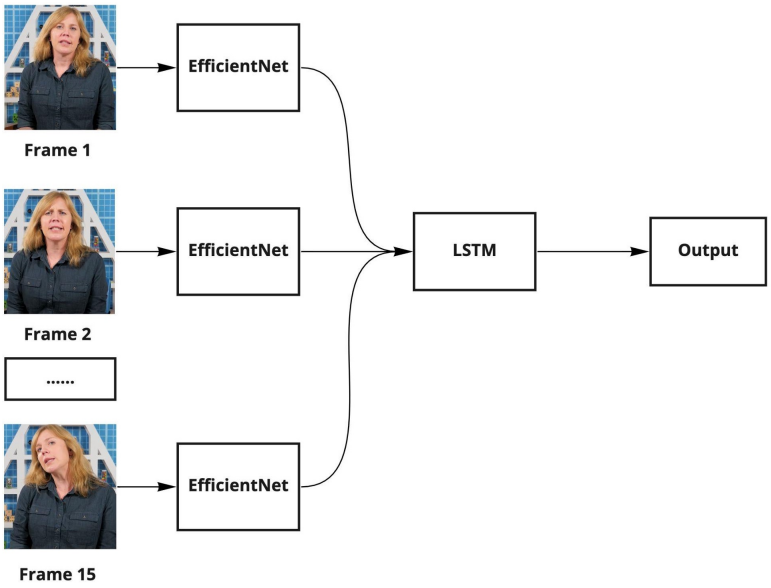

Figure 3 Schematization of Image-Based Deep Learning Model

Table 2 shows the performances of the image-based models. As robustness checks, we also examined the models that use only a single frame instead of multiple frames. Our results showed that EfficientNet B0 achieved the optimal balance between performance and efficiency for this task. We also found that using multiple frames from the video achieved higher accuracy than using single frames. Additionally, we found that overall, the image-based models performed better than the text-based models. Even the MobileNet V2 model, which uses only a single frame, attained higher accuracy than all of the text-based models. This suggests that visual signals presented in images (i.e., what we see) may provide more relevant information for measurement of personality traits versus textual languages in subtitles (i.e., what we read).

4.3. Audio-Based Model

For our audio-based model, we used YAMNet (Plakal and Ellis 2020), which is a deep neural network predicting 521 different audio events in AudioSet. YAMNet first transforms the raw audio

	MobileNet v2, single frame	EfficientNet B0, single frame	EfficientNet B7, single frame	EfficientNet B0, multi frame
Mean Accuracy	0.8996	0.9033	0.9045	0.9052
Extraversion	0.8963	0.9012	0.9007	0.9035
Non-Neuroticism	0.8966	0.8981	0.8989	0.8997
Agreeableness	0.9048	0.9063	0.9067	0.9063
Conscientiousness	0.8991	0.9096	0.9128	0.9110
Openness	0.9012	0.9014	0.9037	0.9053

Table 2 Performance of Image-Based Models

data into a mel spectrogram, and then uses this as the input for the MobileNet V1 architecture in order to derive the predictions for 521 classes of audio. As indicated in our model schematization, shown in Figure 4, we first passed the raw audio through the pretrained YAMNet and then passed the outputs from YAMNet into an LSTM model to capture potential time dependencies. Finally, the output from the LSTM model was passed through two fully connected layers to derive the predictions of each of the five personality traits. For the baseline model, we used pyAudioAnalysis (Giannakopoulos 2015) to extract audio features from raw audio, and then passed it through an LSTM model to derive the final predictions of personality traits.

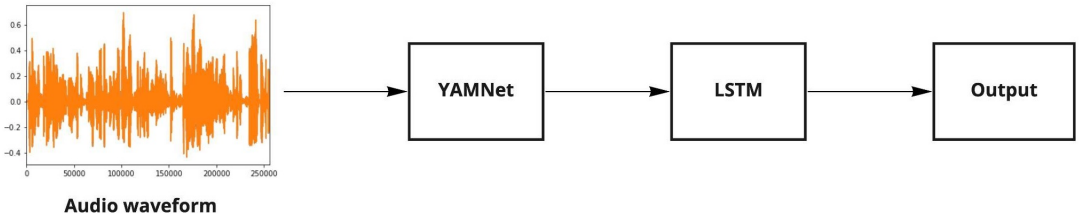

Figure 4 Schematization of Audio-Based Deep Learning Model

Table 3 summarizes the performances of the audio-based models. YAMNet outperformed the baseline model pyAudioAnalysis. Furthermore, we found that overall, the image-based models performed better than the audio-based models as well. This again suggests that visual information contained in images (i.e., what we see) may provide more relevant information for measurement of personality traits compared to auditory or textual information (i.e., what we hear or read), .

	pyAudioAnalysis	YAMNet
Mean Accuracy	0.8891	0.8942
Extraversion	0.8862	0.8922
Non-Neuroticism	0.8853	0.8924
Agreeableness	0.8949	0.8980
Conscientiousness	0.8818	0.8910
Openness	0.8913	0.8966

Table 3 Performances of Audio-Based Models

4.4. Multimodal Model Based on Text, Image, and Audio

Humans interact with the world via different modalities, including language, vision, sound, and smell. Different communicative modalities convey distinct information, which may complement each other in information transmission. For a machine learning model to learn the personality traits from a video, it first needs to combine the messages from multiple communicative modalities. To do this, we built a multimodal deep learning model to predict the personality traits of people in videos trained on the First Impression dataset. Specifically, our model integrates multiple communicative modalities, including visual, linguistic, and acoustic messages, to improve the prediction accuracy of personality traits. Specifically, our model first extracts high-dimensional latent features from image (i.e., video frame), audio (i.e., video soundtrack), and text (i.e., video subtitle) information, separately. It then conducts a late fusion of those features for final personality predictions.

For each communicative modality, we selected the best performing unimodal model (as explained above). Specifically, we used BERT (Devlin et al. 2018) as the feature extractor for the text modality (i.e., subtitles), EfficientNet (Tan and Le 2019) as the feature extractor for the visual modality (i.e., video frames), and YAMNet (Plakal and Ellis 2020) for audio feature extraction. After extracting the latent features from linguistic, visual, and acoustic modalities, we concatenated these features as the final input for the personality traits prediction to produce continuous predictions of the five personality traits ranging from 0 to 1. The architecture of our model is shown in Figure 5.

We summarize the performance comparison of the different models in Table 4. As shown in Table 4, our multimodal model utilizing image, audio, and text data perform better than any of the

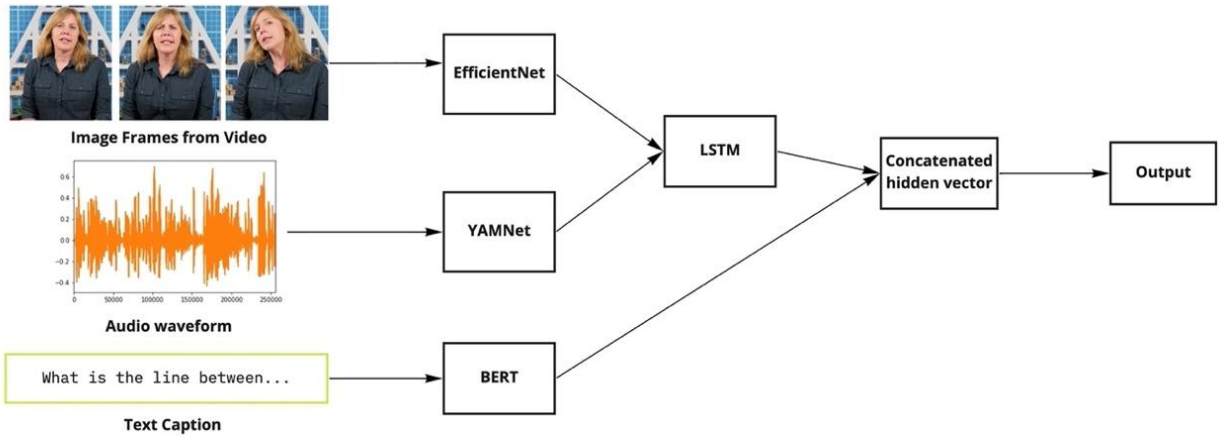

Figure 5 Schematization of Our Proposed Multimodal Deep Learning Model

unimodal models. To test the statistical significance, we conducted t-tests between the multimodal model's performance and each of the unimodal model's performances. The p-values, reported in Table 4, show that our multimodal model's prediction accuracy was significantly higher than that of all of the other models ($p < 0.001$).[1] Figure 6 shows a correlation matrix of the ground truth and different models' predictions. We observe that the image model, the text model, and the audio model showed relatively low correlations with the ground truth labels, whereas our multimodal model achieved the highest correlation with the true labels. This again corroborates that our multimodal model outperformed the other models, and highlights the importance of a multimodal approach that integrates and models multiple communicative modalities in video data.

4.5. Contribution of Each Communicative Modality in Predicting Personality Traits

When people communicate with each other, their personality traits might be reflected in multiple modalities, such as the way they look (i.e., visual messages), talk (i.e., linguistic messages), or sound (i.e., acoustic messages). Since previous personality studies in the IS literature used only textual information to predict personality traits (Yang et al. 2022, Adamopoulos et al. 2018), it is unknown how different modalities may contribute to accurate prediction of personality traits. In particular, what is the importance of each communicative modality in predicting the five personality traits? Given the increasing importance of multimedia data consumption, answering this question has

[1] We conducted both two-sample t-tests and paired t-tests. The results are similar.

	BERT	YAMNet	EfficientNetB0	EfficiencientB0 + YAMNet + BERT
Input	Text	Audio	Image	Text + Audio + Image
Mean Accuracy	0.8899	0.8942	0.9052	0.9105
Extraversion	0.8858	0.8922	0.9035	0.9093
Non-Neuroticism	0.8866	0.8924	0.8997	0.9070
Agreeableness	0.8989	0.8980	0.9063	0.9118
Conscientiousness	0.8865	0.8910	0.9110	0.9146
Openness	0.8915	0.8966	0.9053	0.9094
T-test p-value	< 0.001	< 0.001	< 0.001	

Table 4 Performances of Multimodal Models

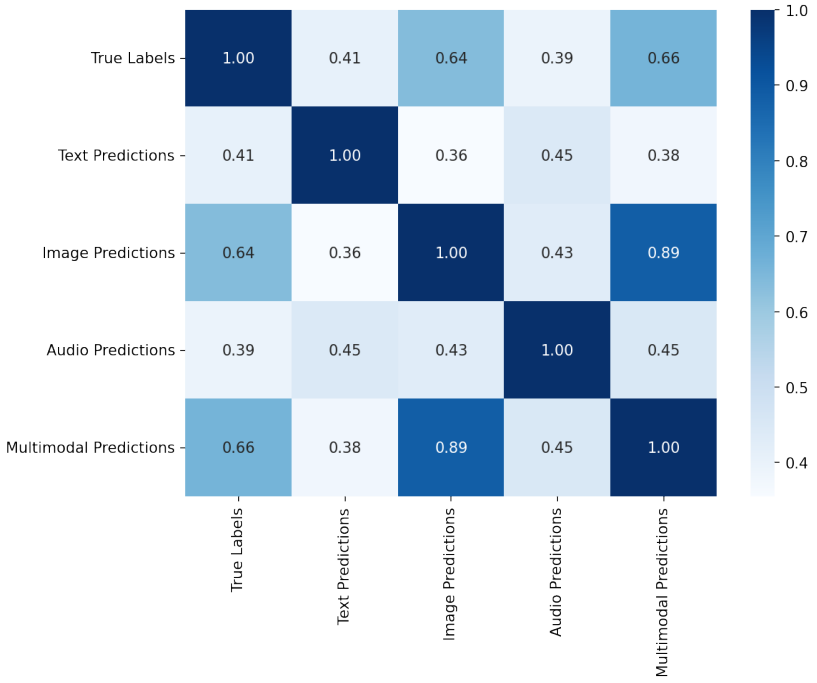

Figure 6 Correlation Matrix of Ground Truth and Different Models' Predictions

significant managerial importance to both content creators and digital platforms. We address this question by further investigating the importance of each modality in our proposed multimodal model.

Instead of conducting an early fusion between image and audio at each second, we performed a late fusion as follows: for both image and audio information, we first used an LSTM model to

capture their trends across time, and then made their own predictions for the personality traits after bypassing two fully connected layers. Similarly, the text modality could also use its information to predict the OCEAN personality traits. In the end, we made a linear combination of the three different modalities' predictions with

$$OCEAN_i = \sum_{j=1}^{3} w_{\{i,j\}} OCEAN_{\{i,j\}} \tag{1}$$

where i denotes the i-th personality trait in the OCEAN personality traits and j denotes the j-th modality. Here, $w_{i,j}$ denotes the weight of modality j for OCEAN personality trait i. We allowed each modality to have different weights on the prediction of each personality trait in order to capture potential heterogeneous effects. The initial value for each weight $w_{i,j}$ was the same (i.e., $\frac{1}{3}$), such that for each personality trait i, $\sum_{j=1}^{3} w_{\{i,j\}} = 1$ at the beginning. We first used the training set to train each modality. Then, after freezing the weights for each modality, we used the validation set to find the weights for each modality and each personality trait. The weights were optimized by minimizing the mean squared error.

Table 5 shows the optimal weights learned from the validation set. As can be seen in the results, even though all three modalities were given the same weight (i.e. $\frac{1}{3}$) before training, their weights differed significantly after learning through the data. In particular, the weights for the image modality were the largest, with an average of 0.78, followed by the weights for the audio modality, with an average of 0.26. The weights for the text modality were the smallest of the three, with an average of -0.02. These results suggest that when predicting personality traits using multimedia data, visual messages may play a dominant role in affecting people's perception of personality traits in the video, whereas the impact of the text modality is relatively negligible in the presence of image and audio information. One implication of this finding is that given the same course materials (e.g., text), one can modify the visual and acoustic presentation (e.g., image and audio) to change the perceived personality traits, which might affect the video popularity as a result.

	Image	Audio	Text
Extraversion	0.87	0.23	-0.08
Non-Neuroticism	0.79	0.23	0.01
Agreeableness	0.69	0.31	-0.01
Conscientiousness	0.92	0.14	0.01
Openness	0.65	0.37	0.00
Average	0.78	0.26	-0.02

Table 5 Contribution of Each Communicative Modality in Predicting Personality Traits

5. Empirical Analysis and Results

Having trained the multimodal deep learning model on the First Impression dataset, we were able to measure the personality traits of the instructors in each educational video in our data. In addition to the prediction performance demonstrated in the previous section, we found that when manually comparing the results for several videos, the predictions from our model are aligned with human judgement. Summary statistics of the five personality traits for our primary dataset of online course videos are reported in Table 6.

Measuring the impact of instructor personality traits is complicated by potential confounding factors in video data. Therefore, in addition to the instructor's personality traits, we extracted a rich set of theory-driven video features for each video in our data to account for potential confounding effects. Specifically, we adopted the video analytics framework in Zhou et al. (2021) to extract basic video properties, instructors' emotions and physical characteristics; as well as visual aesthetic features, which might influence consumers' viewing behavior online. The basic video properties consist of days since video release, video length, speaking rate, average scene length, and sentiment. Days since video release measures the number of days between the video release date and the time when we collected the number of views and likes for that this video. Similar to motion pictures, holding all other things equal, a video that was released earlier will generally accumulate more views or likes than one released later. Video length is the length of a video in minutes. Average speaking rate is the number of spoken words per minute, which is calculated by dividing the number of words in subtitles by the length of the video. Average scene length is calculated by the length

of the video divided by the number of different scenes, where the scenes in the video are identified using an intelligent scene cut detection and video splitting tool named PySceneDetect. Sentiment is derived by applying sentiment analysis to the subtitles of the video, with the values ranging from -1 to 1, where the negative score means negative sentiment, the zero score means neutral sentiment, and the positive score means positive sentiment.

An instructor's emotions and physical characteristics may also affect viewers' perception. We derive these features through Microsoft Azure Face recognition models, and include different emotions (i.e., anger, contempt, disgust, fear, happiness, neutral, sadness, surprise) and physical appearance features (i.e., age, gender, smile, baldness). We found that smile was perfectly collinear with the happiness emotion, which was excluded from our main analysis.

We also extracted visual aesthetic features for each video in our data, including motion features (i.e., foreground motion area, motion magnitude, motion direction) and color features (i.e., warm hue proportion, saturation, brightness, contrast of brightness, clarity). Specifically, foreground motion area, motion magnitude, and motion direction measure the motion characteristics of the video in different ways. Foreground motion area measures the portion of moving pixels in the video. Motion magnitude and motion direction are calculated using the dense optical flow algorithm.

For the color features of the video, warm hue proportion refers to the portion of pixels of warm colors in a frame. Saturation is the average of intensity of color in the video. Brightness refers to the average intensity values of all pixels in the video, whereas the contrast of brightness is defined as the standard deviation of intensity values of all pixels. Clarity is the portion of pixels with sufficient brightness in each frame. These visual aesthetic features were calculated for each frame in the video. We then aggregated the values to the video level by taking the average values across the entire video. Table 6 summarizes the main variables used in the analysis and shows the corresponding descriptive statistics.

5.1. Double Machine Learning Framework

Using this rich set of theory-driven video features as control variables, we use the double machine learning (DML) framework to estimate the impact of the five personality traits on the success of

Variable	Description	Mean	SD	Min	Max
Dependent Variables:					
Views_count	Number of views of a video	383,165.433	815,461.435	580.000	11,904,048.000
Likes_count	Number of likes of a video	6,914.394	10,876.640	5.000	165,754.000
Main-Effect Variables:					
Extraversion	Level of extraversion in the personality of an instructor	0.4267	0.0631	0.2548	0.6458
Non-Neuroticism	Level of neuroticism in the personality of an instructor	0.5382	0.0538	0.3373	0.6747
Agreeableness	Level of agreeableness in the personality of an instructor	0.5746	0.0478	0.4308	0.7435
Conscientiousness	Level of conscientiousness in the personality of an instructor	0.6238	0.0567	0.3927	0.7594
Openness	Level of openness in the personality of an instructor	0.5410	0.0585	0.3591	0.7198
Control Variables:					
Day since release	Number of days since first release	1,443.086	1,017.666	1.000	4,614.000
Video length	The length of the video measured in minutes	23.038	25.513	0.449	184.106
Speaking rate	The number of spoken words per minute in the video	232.230	133.777	12.706	669.602
Sentiment	The average sentiment of sentences in the subtitle file	0.801	0.561	-1.000	1.000
Average scene length	The average length of a scene in the video, measured in minutes computed based on intelligent scene cut detection	1.490	4.691	0.055	70.928
Warm hue proportion	The portion of pixels in warm colors (e.g., yellow, red) in a frame	0.549	0.254	0.029	1.000
Saturation	Average saturation across all pixels in a frame	0.352	0.165	0.001	0.860
Brightness	Average intensity across all pixels in a frame	0.473	0.152	0.044	0.978
Contrast	The standard deviation of pixel intensity values across the whole frame	0.202	0.054	0.072	0.467
Clarity	The portion of pixels with sufficient intensity in a frame	0.968	0.071	0.196	1.000
Foreground motion	The average percentage of foreground motion area in the video, computed based on foreground/background segmentation	0.195	0.118	0.000	0.626
Motion magnitude	Average motion magnitude measured in pixels, computed based on dense optical flow	0.440	0.299	0.001	4.016
Motion direction	Average motion direction measured in degrees, computed based on dense optical flow	3.085	0.120	1.650	3.507
Anger	Measure of Anger, computed using a pre-trained deep learning model to detect the emotions of the instructor in each video	0.007	0.023	0.000	0.527
Contempt	Measure of Contempt computed using a pre-trained deep learning model to detect the emotions of the instructor in each video	0.009	0.017	0.000	0.235
Disgust	Measure of Disgust, computed using a pre-trained deep learning model to detect the emotions of the instructor in each video	0.003	0.013	0.000	0.362
Fear	Measure of Fear, computed using a pre-trained deep learning model to detect the emotions of the instructor in each video	0.002	0.007	0.000	0.179
Happiness	Measure of Happiness, computed using a pre-trained deep learning model to detect the emotions of the instructor in each video	0.216	0.229	0.000	1.000
Neutral	Measure of Neutral, computed using a pre-trained deep learning model to detect the emotions of the instructor in each video	0.647	0.234	0.000	1.000
Sadness	Measure of Sadness, computed using a pre-trained deep learning model to detect the emotions of the instructor in each video	0.028	0.062	0.000	0.692
Surprise	Measure of Surprise, computed using a pre-trained deep learning model to detect the emotions of the instructor in each video	0.088	0.098	0.000	0.668
Age	Measure of Age, computed using a pre-trained deep learning model for face detection and classification	41.395	10.486	9.400	73.000
Gender	Measure of Gender, computed using a pre-trained deep learning model for face detection and classification	0.834	0.372	0.000	1.000
Smile	Measure of Smile, computed using a pre-trained deep learning model for face detection and classification	0.216	0.229	0.000	1.000
Baldness	Measure of Baldness, computed using a pre-trained deep learning model for face detection and classification	0.175	0.231	0.000	0.989

Table 6 Main Variables and Descriptive Statistics

online educational videos as measured by the number of likes or views. DML provides a general framework to derive consistent estimates of a low-dimensional parameter of interest θ_0 when there exists a high-dimensional nuisance parameter η_0 (Chernozhukov et al. 2018). DML protects against bias due to model mis-specification, avoids reliance on unrealistic parametric distributions, and reduces the curse of dimensionality commonly faced in the presence of big data. Consider the partially linear regression model discussed in Chernozhukov et al. (2018):

$$\begin{aligned} Y &= D\theta_0 + g_0(X) + U, E[U|X, D] = 0 \\ D &= m_0(X) + V, E[V|X] = 0 \end{aligned} \tag{2}$$

where Y denotes the outcome variable, D the policy or treatment variable of interest, vector X the possible confounders, and U and V the error terms. Our parameter of interest θ_0 is contained in the first equation. The second equation models the relationship between treatment D and confounders X. We are not interested in the specific forms of g_0 or m_0, so they are denoted as nuisance parameters $\eta = (g_0, m_0)$. The nuisance parameters η are allowed to have complex forms, so as to better capture the potential high-dimensional confounders X, or the complex, non-linear relationships of the confounders X. Compared with simple linear models, this enables better modeling of the relationships between the confounders X and the outcome Y, to avoid possible model mis-specification.

A naive approach is to use machine learning algorithms to directly estimate the first equation in Eq.2. However, Chernozhukov et al. (2018) shows that this naive estimator would fail to converge in the $N^{-\frac{1}{2}}$ rate, where N is the number of samples. More specifically, this bias comes from the possible over-fitting and regularization problem in machine learning models. DML uses orthogonalization and sample-splitting to remove these biases to get an $N^{-\frac{1}{2}}$ consistent estimator under mild assumptions. More specifically, the general DML framework is as follows:

1. Use any machine learning method to estimate l_0 and m_0 and get the residuals: $l_0(x) = E(Y|x)$, so $\hat{W} = Y - \hat{l_0}(X)$; $m_0(x) = E(U|x)$, so $\hat{V} = D - \hat{m_0}(X)$. Since the effect of confounders X on outcome Y and on treatment D has been partialled out, here $\hat{W}$ and $\hat{V}$ are orthogonal.

2. Regress $\hat{W}$ on $\hat{V}$ using linear regression to get an estimate $\check{\theta}_0$.

3. Cross-fitting. We randomly split the sample into K folds. Let I_k and I_k^c where $k \in 1, \cdots, K$ denote each fold and its complement. Then for each $k \in 1, \cdots, K$, we use data I_k to estimate Step 1, and use data I_k^c to estimate Step 2 so as to get $\check{\theta}_{0,k}$. Finally, we take the average of K $\check{\theta}_{0,k}$ to get the final estimate $\check{\theta}_0$.

Our setting is a bit more complicated than the basic model in Eq.2, since we have multiple treatment variables (i.e., five different personality traits). Therefore, our problem falls into the setting to conduct simultaneous inference for multiple treatment variables. Now, consider the case where there are p_1 number of treatments $D_1, D_2, \cdots, D_{p_1}$ and the corresponding variables of interest are $\theta_1, \theta_2, \cdots, \theta_{p_1}$. Then, this simultaneous inference is performed by iteratively performing DML on each variable of interest. More specifically, each parameter of interest θ_j where $j \in 1, \cdots, p_1$, is derived as follows. First, the main equation is modified as

$$Y = D_j\theta_j + g_{0,j}(X_j) + U_j, E[U_j|X_j, D_j] = 0 \tag{3}$$

where $X_j = [X, D_1, D_2, \cdots, D_{j-1}, D_{j+1}, \cdots, D_{p_1}]$. Similarly, the equation for modelling the relationship between treatment D_j and confounders is modified as

$$D_j = m_{0,j}(X_j) + V_j, E[V_j|X_j] = 0 \tag{4}$$

In other words, when estimating the effect of D_j, we add the other four treatment variables as confounders so as to capture their relationships with D_j and Y. Correspondingly, the nuisance parameters when estimating θ_j is $\eta_j = (g_{0,j}, m_{0,j})$. For more details regarding simultaneous inference for multiple treatment variables, please refer to Belloni et al. (2018).

In out setting, Y denotes the outcome variable, the number of likes or views for each video. We performed log-transformation on the outcome variables to reduce the skewness of the variables. X denotes the possible confounders (i.e., various video features), and $D_1, D_2, \cdots, D_5$ denote the treatment variables (i.e., the five personality traits). To estimate nuisance parameters $g_{0,j}$ and

$m_{0,j}$, we used XGBoost (Chen and Guestrin 2016) which is a scalable implementation of gradient-boosted decision trees. This is an ensemble model that fits a series of decision trees based on the previous residuals so as to minimize the specific loss function such as mean squared loss. It has been observed in research to win many machine learning challenges with different applications (Fu et al. 2021). We estimated the main effects of various instructor personality traits using the DoubleML package (Bach et al. 2022) to get the estimates for $\theta_1, \theta_2, \cdots, \theta_5$ respectively.[2]

5.2. Results and Discussion

We next present our main estimation results for the effect of instructor personality traits on the success of online educational videos. We repeated the DML analysis for two dependent variables: the natural logarithm of the number of views and natural logarithm of the number of likes for each video. In addition to an extensive set of control variables extracted from unstructured video data, we also include channel-level fixed effects to account for the inherent differences between different educational channels on YouTube. Our main results are summarized in Table 7. As can be seen in columns (1) and (3), there is a positive and statistically significant effect of an instructor's level of extraversion on the popularity of online educational videos, whereas the effect of the level of openness was negative and statistically significant. Moreover, the results for using two different dependent variables were similar in both magnitude and significance level. These effects were found to be robust when we further included channel-level fixed effects, as shown in columns (2) and (4).

In addition to the main effects of the instructor's latent personality traits, we further investigate how these effects may be moderated by the instructor's observable characteristics such as age and gender. Previous studies in psychology have shown that age is an moderator in different settings. For example, Buecker et al. (2020) found that the relation between loneliness and personality is moderated by age, and Mammadov (2022) found that the strength of associations between student personality traits and academic performance is moderated by student age. In our setting, the age

[2] DML allows to use any machine learning algorithm to estimate the nuisance parameters. We have also tried different algorithms such as Random Forest which show similar results.

	Dependent variables:			
	log(views_count)		log(likes_count)	
	(1)	(2)	(3)	(4)
Extraversion	2.350***	1.675***	2.644***	1.696***
	(0.543)	(0.500)	(0.557)	(0.510)
Non-Neuroticism	-0.199	-0.161	-0.353	-0.307
	(0.709)	(0.656)	(0.757)	(0.683)
Agreeableness	-0.445	-0.531	-0.041	-0.128
	(0.669)	(0.629)	(0.709)	(0.649)
Conscientiousness	0.328	0.539	0.671	0.523
	(0.578)	(0.548)	(0.607)	(0.561)
Openness	-1.538**	-1.115*	-1.829**	-1.142*
	(0.575)	(0.551)	(0.610)	(0.562)
N	6,090	6,090	6,090	6,090
Channel-level FE	No	Yes	No	Yes
Note:			$^*p < 0.05$; $^*p < 0.01$; $^*p < 0.001$	

Table 7 Double Machine Learning (DML) Estimation Results on the Effects of Personality Traits

of an instructor might also be a potential moderator, because instructors of different ages may give different impressions in videos. In order to analyze the potential moderating effect of instructor age, we first split the sample into two groups (i.e., old versus young) using the median age of the entire sample as the threshold. After this sample splitting, we first compare the differences in instructor personality traits between these two groups. The summary statistics for each group are presented in Table 8. Interestingly, we found that the average values of the instructor's personality traits were statistically higher in the young group than those in the old group, suggesting that younger instructors tend to display stronger personalities in videos.

We next estimate the effects of instructor personality traits using the same DML framework on the old and young groups, respectively. Table 9 shows the results for each group. We find that the effects of personality traits are heterogeneous between the groups. Specifically, for the old group, there was a positive and statistically significant effect of the instructor's level of extraversion on the popularity of online educational videos, as shown in columns (1) and (3). Even though young instructors tended to show stronger personality traits, and to seem more extroverted in the video compared with the old instructors, the positive effect of extraversion found in Table 7 was mainly

	Old		Young		T-test
Personality	Mean	SD	Mean	SD	p-value
Extraversion	0.411	0.060	0.443	0.062	< 0.001
Non-Neuroticism	0.529	0.055	0.547	0.051	< 0.001
Agreeableness	0.570	0.047	0.579	0.048	< 0.001
Conscientiousness	0.622	0.057	0.626	0.057	< 0.01
Openness	0.526	0.057	0.556	0.056	< 0.001
N	3,055		3,035		

Table 8 Comparison of Personality Traits between Old and Young Instructors

driven by the old group. On the contrary, for the young group, there was a negative and statistically significant effect of the instructor's level of openness on the popularity of online educational videos, as shown in columns (2) and (4), whereas such effect was not significant for the old group. This suggests that the negative effect of openness found in Table 7 was mainly driven by the young group.

	Dependent variable:			
	log(views_count)		log(likes_count)	
	(1)	(2)	(3)	(4)
Extraversion	2.051**	0.550	2.253**	0.379
	(0.779)	(0.636)	(0.803)	(0.639)
Non-Neuroticism	-1.012	1.311	-1.374	0.909
	(0.969)	(0.875)	(1.026)	(0.874)
Agreeableness	-0.186	-0.694	0.411	-0.830
	(0.895)	(0.882)	(0.923)	(0.902)
Conscientiousness	-0.169	0.910	0.187	1.165
	(0.777)	(0.726)	(0.809)	(0.737)
Openness	-0.430	-2.513***	-0.742	-2.257**
	(0.825)	(0.697)	(0.842)	(0.713)
Group	Old	Young	Old	Young
N	3,055	3,035	3,055	3,035
Note:		$^*p < 0.05$; $^*p < 0.01$; $^*p < 0.001$		

Table 9 Different Effects of Personality Traits for Old vs. Young Instructors

Next, we further explore gender as a moderator of the effects of instructors' personality traits on the popularity of their online educational videos. Prior research has suggested that there are

many ways gender can play a role in our analysis. For example, Nguyen et al. (2005) found that students' personality traits can predict their academic performance, and that this personality-academic performance relationship is moderated by student gender. In a similar vein, Nahyun and Hana (2011) determined that gender also moderates the effect of students' personality traits on their information competency, and Asghari et al. (2013) noted that gender moderates the relationship between students' personality traits and examination anxiety. In our setting, we focused on the instructor's personality traits and examined whether the effects of those personality traits differed across male and female instructors.

Table 10 presents the summary statistics for these two groups, which show significant differences in personality traits between male and female instructors. We used the same DML framework to estimate the effects of instructor personality traits for male and female instructors, respectively. The estimation results for each group are presented in Table 11, which shows differences between male and female instructors. In particular, for male instructors, there was a positive and statistically significant effect of the instructor's level of extraversion on the popularity of online educational videos, as shown in columns (1) and (3), whereas such positive effect of extraversion was not significant for female instructors, as shown in columns (2) and (4). Similarly, for the male group, there was a negative and statistically significant effect of the instructor's level of openness on the popularity of online educational videos, whereas such effect was not significant for the female instructors.

	Male		Female		T-test
Personality	Mean	SD	Mean	SD	p-value
Extraversion	0.420	0.060	0.461	0.068	< 0.001
Non-Neuroticism	0.537	0.054	0.542	0.053	< 0.05
Agreeableness	0.576	0.047	0.567	0.051	< 0.001
Conscientiousness	0.626	0.057	0.613	0.056	< 0.001
Openness	0.535	0.056	0.573	0.060	< 0.001
N	5,078		1,012		

Table 10 Comparison of Personality Traits between Male and Female Instructors

	Dependent variables:			
	log(views_count)		log(likes_count)	
	(1)	(2)	(3)	(4)
Extraversion	1.976***	1.508	1.958***	0.855
	(0.576)	(1.153)	(0.588)	(1.087)
Non-Neuroticism	0.006	-0.019	-0.033	-1.239
	(0.733)	(1.498)	(0.763)	(1.482)
Agreeableness	-0.888	1.428	-0.819	1.728
	(0.703)	(1.437)	(0.728)	(1.492)
Conscientiousness	0.656	-0.726	0.781	-0.369
	(0.595)	(1.371)	(0.616)	(1.389)
Openness	-1.498*	-1.650	-1.605*	-0.939
	(0.610)	(1.353)	(0.630)	(1.358)
Group	Male	Female	Male	Female
N	5,078	1,012	5,078	1,012
Note:				$^*p < 0.05$; $^*p < 0.01$; $^*p < 0.001$

Table 11 Different Effects of Personality Traits for Male vs. Female Instructors

Lastly, we further investigate the heterogeneity across different course topics. Specifically, we first classify each course video into one of the two groups: (1) science, technology, engineering, and mathematics (STEM) versus (2) non-STEM. We then explore whether the effects of instructors' personality traits on the popularity of their online educational videos differed across the topics.

Table 12 presents the summary statistics for these two groups. We repeated the DML analysis to estimate the effects of instructor personality traits for these two groups, respectively. The estimation results for each group are presented in Table 13, which shows some heterogeneous effects for STEM versus non-STEM courses. In particular, for both STEM and non-STEM courses, there was a positive and statistically significant effect of the instructor's level of extraversion on the popularity of online educational videos. However, for STEM courses, there was a negative and statistically significant effect of the instructor's level of openness on the popularity of online educational videos, as shown in columns (1) and (3), whereas such negative effect of openness was not significant for non-STEM courses, as shown in columns (2) and (4).

	STEM		non-STEM		T-test
Personality	Mean	SD	Mean	SD	p-value
Extraversion	0.422	0.060	0.439	0.070	< 0.001
Non-Neuroticism	0.537	0.053	0.540	0.054	0.08
Agreeableness	0.574	0.047	0.575	0.049	0.49
Conscientiousness	0.625	0.056	0.619	0.057	< 0.001
Openness	0.539	0.057	0.547	0.062	< 0.001
N	4,472		1,602		

Table 12 Comparison of Personality Traits between STEM vs. non-STEM Instructors

	Dependent variables:			
	log(views_count)		log(likes_count)	
	(1)	(2)	(3)	(4)
Extraversion	1.627**	1.660	1.635**	2.451**
	(0.570)	(0.894)	(0.583)	(0.896)
Non-Neuroticism	-0.340	-0.753	-0.879	-1.393
	(0.769)	(1.075)	(0.799)	(1.075)
Agreeableness	0.601	-1.637	0.537	-0.698
	(0.747)	(1.058)	(0.748)	(1.075)
Conscientiousness	0.578	0.637	0.535	0.469
	(0.647)	(0.929)	(0.654)	(0.950)
Openness	-1.891**	-0.091	-1.539*	-0.671
	(0.632)	(1.004)	(0.643)	(0.988)
Group	STEM	non-STEM	STEM	non-STEM
N	4,472	1,602	4,472	1,602
Note:				$^*p < 0.05$; $^*p < 0.01$; $^*p < 0.001$

Table 13 Different Effects of Personality Traits for STEM vs. non-STEM Instructors

6. Conclusions

Education is one of the most important industries for the global economy and for global social progress. Information technology has brought significant changes to this industry and transformed the creation, distribution, and consumption of educational content. With the rapid development of technology, and the disruption caused by the COVID-19 pandemic, the online education market has been growing substantially over the past few years. Despite its economic and societal importance, there is as yet little empirical research on the determinants of the success of educational

materials in the online environment. At the same time, the abundance of online educational content provides, to both online education platforms and researchers, the opportunity to directly observe more fine-grained data and gain deeper insights into this industry. In this study, tapping into the idea that personality characteristics may affect individual behaviors, and leveraging deep learning and video-mining techniques, we extend the literature by providing the first empirical evidence on the relationship between instructors' latent personality traits and their online course videos' performance. In essence, this paper takes a significant step towards the goal of content engineering for improved online education effectiveness. In particular, drawing on theories that are rooted in psychology and the social sciences, we examine at a granular level how specific personality traits of the instructor affect the popularity of online course videos and how such effects might differ across different instructors.

In addition, our study contributes to the burgeoning business analytics research by synergistically using personality theories and data analytics. The rise of unstructured data is reshaping business practices in many settings, and has attracted the attention from researchers in many domains. More and more studies in the Information Systems literature are using machine learning techniques to extract insights from unstructured data such as texts or images. Our research expands the literature's unimodal scope to a multimodal one by integrating and modeling different types of unstructured data (e.g., text, image, audio) from the multiple communicative modalities in video content. Using a unique large-scale video dataset, we developed a multimodal deep learning model to predict the latent personality traits, and demonstrated its superior performance over commonly used unimodal text-based predictions. This model, by using increasingly available video data and suitable analytics techniques to measure constructs that were previously costly to measure, complements the traditional, labor intensive, social science method of using surveys or interviews to assess personality. By using a multimodal predictive model approach, information systems researchers can assess an instructor's personality automatically and instantaneously. This adds a new methodological approach from the design science perspective (Gregor and Hevner

2013), and provides practical and actionable implications for online video platforms. As digital video consumption has become an essential part of our daily lives, our findings and methodologies have broad implications for various business domains such as online education, video marketing, and livestreaming e-commerce.

Whereas this paper takes an important step towards understanding the growing online education market as well as video analytics in IS research, we acknowledge several limitations to our research. First, in addition to users' likes and views on online education platforms, future research should study how an instructor's latent personality traits affect both student engagement and student learning outcomes measured by assignment or exam grades. Second, future studies may also expand the analysis to a more fine-grained level such as by using eye-tracking analysis in a laboratory environment to better understand detailed user behaviors during the video watching process. Third, given the increasing ubiquity of online video consumption in domains ranging widely from entertainment to education, future research can build on this paper to examine the heterogeneous effects of personality traits on online video platforms in various fields such as video advertising and influencer marketing.

References

Adamopoulos P (2013) What makes a great MOOC? An interdisciplinary analysis of student retention in online courses. *34th International Conference on Information Systems: ICIS 2013* (Association for Information Systems).

Adamopoulos P, Ghose A, Todri V (2018) The impact of user personality traits on word of mouth: Text-mining social media platforms. *Information Systems Research* 29(3):612–640.

Ahmad F, Abbasi A, Li J, Dobolyi DG, Netemeyer RG, Clifford GD, Chen H (2020) A deep learning architecture for psychometric natural language processing. *ACM Transactions on Information Systems (TOIS)* 38(1):1–29.

AmericanPsychologicalAssociation (2022) Personality. URL `https://www.apa.org/topics/personality`.

Antoncic B, Bratkovic Kregar T, Singh G, DeNoble AF (2015) The big five personality–entrepreneurship relationship: Evidence from slovenia. *Journal of Small Business Management* 53(3):819–841.

Asghari A, bte Elias H, bte Baba M, et al. (2013) Personality traits and examination anxiety: Moderating role of gender. *Alberta Journal of Educational Research* 59(1):45–54.

Bach P, Chernozhukov V, Kurz MS, Spindler M (2022) DoubleML – An object-oriented implementation of double machine learning in Python. *Journal of Machine Learning Research* 23(53):1–6, URL `http://jmlr.org/papers/v23/21-0862.html`.

Back MD, Stopfer JM, Vazire S, Gaddis S, Schmukle SC, Egloff B, Gosling SD (2010) Facebook profiles reflect actual personality, not self-idealization. *Psychological Science* 21(3):372–374.

Belloni A, Chernozhukov V, Chetverikov D, Wei Y (2018) Uniformly valid post-regularization confidence regions for many functional parameters in z-estimation framework. *Annals of Statistics* 46(6B):3643.

Bradley RA, Terry ME (1952) Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika* 39(3/4):324–345.

Buecker S, Maes M, Denissen JJ, Luhmann M (2020) Loneliness and the big five personality traits: A meta–analysis. *European Journal of Personality* 34(1):8–28.

Chen G, Davis D, Hauff C, Houben GJ (2016) On the impact of personality in massive open online learning. *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization*, 121–130.

Chen J, Haber E, Kang R, Hsieh G, Mahmud J (2015) Making use of derived personality: The case of social media ad targeting. *Ninth International AAAI Conference on Web and Social Media*.

Chen T, Guestrin C (2016) Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.

Chernozhukov V, Chetverikov D, Demirer M, Duflo E, Hansen C, Newey W, Robins J (2018) Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal* .

Costa PT, McCrae RR, Dye DA (1991) Facet scales for agreeableness and conscientiousness: A revision of the neo personality inventory. *Personality and Individual Differences* 12(9):887–898.

Dellarocas C, Van Alstyne M (2013) Money models for moocs. *Communications of the ACM* 56(8):25–28.

Devaraj S, Easley RF, Crant JM (2008) Research note—how does personality matter? relating the five-factor model to technology acceptance and use. *Information Systems Research* 19(1):93–105.

Devlin J, Chang MW, Lee K, Toutanova K (2018) Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* .

Fu R, Huang Y, Singh PV (2021) Crowds, lending, machine, and bias. *Information Systems Research* 32(1):72–92.

Gerber AS, Huber GA, Doherty D, Dowling CM (2011) The big five personality traits in the political arena. *Annual Review of Political Science* 14:265–287.

Giannakopoulos T (2015) PyAudioAnalysis: An open-source python library for audio signal analysis. *PloS one* 10(12).

Giluk TL, Postlethwaite BE (2015) Big five personality and academic dishonesty: A meta-analytic review. *Personality and Individual Differences* 72:59–67.

Goldberg LR (1990) An alternative "description of personality": The big-five factor structure. *Journal of Personality and Social Psychology* 59(6):1216.

Gregor S, Hevner AR (2013) Positioning and presenting design science research for maximum impact. *MIS Quarterly* 337–355.

He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.

Huang N, Zhang J, Burtch G, Li X, Chen P (2021) Combating procrastination on massive online open courses via optimal calls to action. *Information Systems Research* .

Investopedia (2019) What country spends the most on education? URL `https://www.investopedia.com/ask/answers/020915/what-country-spends-most-education.asp#:~:text=In%20terms%20of%20the%20percentage,Kingdom%2C%20Colombia%2C%20and%20Chile.`

InVideo (2022) 135 video marketing statistics you can't ignore in 2022. URL `https://invideo.io/blog/video-marketing-statistics/`.

John OP, Srivastava S, et al. (1999) The big five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of Personality: Theory and Research* 2(1999):102–138.

Jordan K (2015) MOOC completion rates: The data. URL `http://www.katyjordan.com/MOOCproject.html`.

Kim LE, Jörg V, Klassen RM (2019) A meta-analysis of the effects of teacher personality on teacher effectiveness and burnout. *Educational Psychology Review* 31(1):163–195.

Komarraju M, Karau SJ, Schmeck RR (2009) Role of the big five personality traits in predicting college students' academic motivation and achievement. *Learning and Individual Differences* 19(1):47–52.

Leung ACM, Santhanam R, Kwok RCW, Yue WT (2022) Could gamification designs enhance online learning through personalization? lessons from a field experiment. *Information Systems Research* .

Liu AX, Li Y, Xu SX (2021) Assessing the unacquainted: Inferred reviewer personality and review helpfulness. *MIS Quarterly* 45(3).

Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, Levy O, Lewis M, Zettlemoyer L, Stoyanov V (2019) Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692* .

Liu Z, Wang Y, Mahmud J, Akkiraju R, Schoudt J, Xu A, Donovan B (2016) To buy or not to buy? understanding the role of personality traits in predicting consumer behaviors. *International Conference on Social Informatics*, 337–346 (Springer).

Mammadov S (2022) Big five personality traits and academic performance: A meta-analysis. *Journal of Personality* 90(2):222–255.

Martincin KM, Stead GB (2015) Five-factor model and difficulties in career decision making: A meta-analysis. *Journal of Career Assessment* 23(1):3–19.

Matcha W, Gašević D, Jovanović J, Uzir NA, Oliver CW, Murray A, Gasevic D (2020) Analytics of learning strategies: The association with the personality traits. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 151–160.

Matthews G, Deary IJ, Whiteman MC (2003) *Personality traits* (Cambridge University Press).

Nahyun K, Hana S (2011) Personality, traits, gender and information competency among college students. *Malaysian Journal of Library & Information Science* 16(1):87–107.

Nguyen N, Allen LC, Fraccastoro K (2005) Personality predicts academic performance: Exploring the moderating role of gender. *Journal of Higher Education Policy and Management* 27(1):105–117.

Perry E (2019) 2020 video marketing and statistics: What brands need to know. URL `https://socialmediaweek.org/blog/2019/10/2020-video-marketing-and-statistics-what-brands-need-to-know/`.

Plakal M, Ellis D (2020) Sound classification with yamnet.

Ponce-López V, Chen B, Oliu M, Corneanu C, Clapés A, Guyon I, Baró X, Escalante HJ, Escalera S (2016) Chalearn lap 2016: First round challenge on first impressions-dataset and results. *European Conference on Computer Vision*, 400–418 (Springer).

Richardson M, Abraham C, Bond R (2012) Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychological Bulletin* 138(2):353.

Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC (2018) Mobilenetv2: Inverted residuals and linear bottlenecks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4510–4520.

Statista (2022) E-learning: Global market size by segment. URL `https://www.statista.com/statistics/1130331/e-learning-market-size-segment-worldwide/`.

Stone M (2018) Education and marketing: Decision making, spending, and consumption. URL `https://www.ama.org/2018/09/21/journal-of-marketing-research-special-issue-education-and-marketing/`.

Tan M, Le Q (2019) Efficientnet: Rethinking model scaling for convolutional neural networks. *International Conference on Machine Learning*, 6105–6114 (PMLR).

TechSmith (2022) Why video is important: What you need to know. URL `https://www.techsmith.com/blog/why-video-is-important/`.

Terwiesch C, Ulrich KT (2014) Will video kill the classroom star? The threat and opportunity of massively open on-line courses for full-time mba programs. *SSRN 2467557* .

UNESCO (2022) Education: From disruption to recovery. URL `https://en.unesco.org/covid19/educationresponse`.

UnitedNations (2020) Policy brief: Education during covid-19 and beyond. *https://techjury.net/blog/elearning-statistics/* .

Venkatesh V, Sykes TA, Venkatraman S (2014) Understanding e-government portal use in rural india: role of demographic and personality characteristics. *Information Systems Journal* 24(3):249–269.

Wojcicki S (2018) My five priorities for creators in 2018. URL `https://blog.youtube/inside-youtube/my-five-priorities-for-creators-in-2018_1/`.

Yang K, Lau RY, Abbasi A (2022) Getting personal: A deep learning artifact for text-based measurement of personality. *Information Systems Research* .

Yang Z, Dai Z, Yang Y, Carbonell J, Salakhutdinov RR, Le QV (2019) Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in Neural Information Processing Systems* 32.

Zhang DJ, Allon G, Van Mieghem JA (2017) Does social interaction improve learning outcomes? evidence from field experiments on massive open online courses. *Manufacturing & Service Operations Management* 19(3):347–367.

Zhao K, Hu Y, Hong Y, Westland JC (2019) Understanding characteristics of popular streamers on live streaming platforms: Evidence from twitch. tv. *Journal of the Association for Information Systems* .

Zhou M, Chen GH, Ferreira P, Smith MD (2021) Consumer behavior in the online classroom: Using video analytics and machine learning to understand the consumption of video courseware. *Journal of Marketing Research* 58(6):1079–1100.

Appendix A: Details on Bidirectional Encoder Representations from Transformers (BERT)

As discussed in the main text, when extracting information on instructors' personality from textual data, we chose to fine-tune BERT on the First Impression Dataset. BERT first conducts pre-training on massive unlabelled text to learn the joint latent representations conditioned on both the left and right contexts. More specially, it is trained on two tasks: Masked Language Model (Masked LM) and Next Sentence Prediction (NSP). In Masked LM, some words in a given sentence are randomly masked, and the model is asked to predict the masked words. For the NSP task, a pair of masked Sentences A and B serve as the input, and the model is asked to predict whether Sentence B is the next sentence after Sentence A. Then, after pre-training, with one additional output layer, it is fine-tuned by labeled text data on specific tasks, such as sentiment analysis and question answering. Our task (personality prediction) is similar to sentiment analysis, and so we just added two fully connected layers and, lastly, a sigmoid activation function to render the output on the $[0, 1]$ scale, which corresponds to the range of the five personality traits.

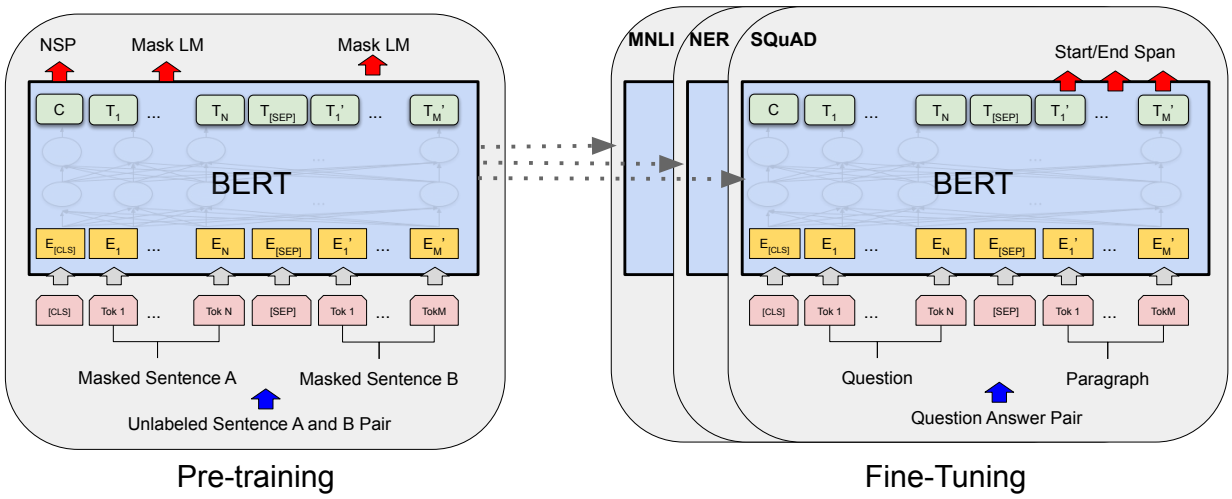

Figure 7 Overall pre-training and fine-tuning procedures for BERT (Devlin et al. 2018)

Appendix B: Details on EfficientNet

As discussed in the main text, when extracting information on instructors' personality from image data, we chose to fine-tune EfficientNet on the First Impression Dataset. EfficientNet is a series of convolutional neural network structures that optimize three major factors in scaling convoluted neural networks: width, depth, and resolution, jointly. Intuitively, increasing any of those three factors would boost performance, but

performing a grid search would be inefficient. EfficientNet, as its name suggests, optimizes them jointly and then propose a family of network models. This joint-optimization intuition can be represented as in Figure 8.

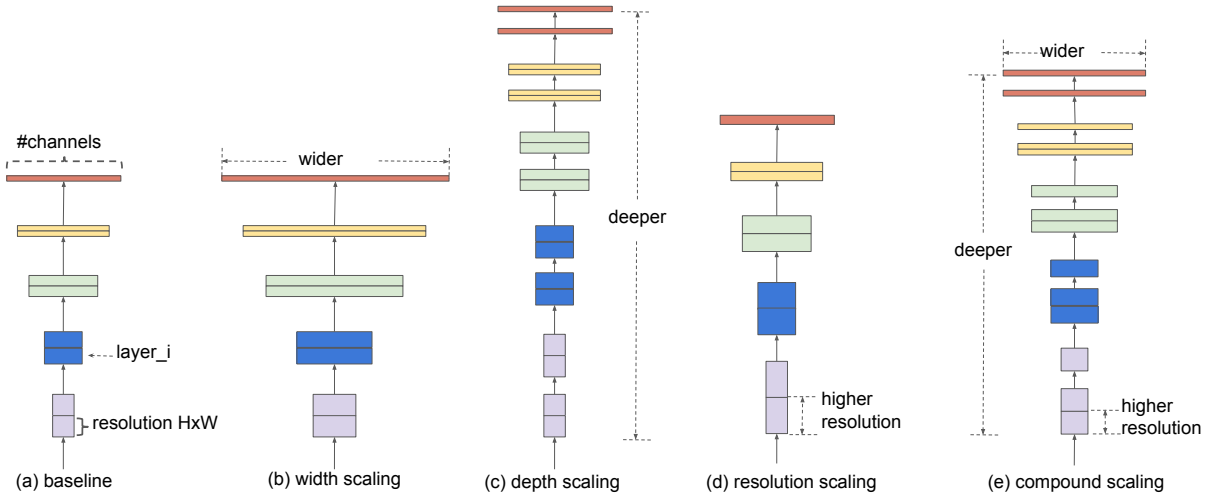

Figure 8 EfficientNet Model Scaling (Tan and Le 2019)

Here (a) is a baseline network example; (b)-(d) are conventional scaling that only increases one dimension of network width, depth, or resolution; (e) is EfficientNet's proposed compound scaling method that jointly scales all three dimensions according to a fixed ratio (Tan and Le 2019).

The baseline network Efficient B0, which is also our model of choice, has the structure shown in Table 14. Its main building block is mobile inverted bottleneck MBConv introduced in Sandler et al. (2018), which makes it more memory efficient than the traditional residual block (He et al. 2016). Our model adds two fully connected layers, and lastly, a sigmoid activation function to render the output on the $[0, 1]$ scale, which corresponds to the range of the five personality traits.

Table 14 **EfficientNet-B0 baseline network (Table extracted from Tan and Le (2019)). Each row represents a stage i with $\hat{L}_i$ layers, with input resolution $\langle \hat{H}_i, \hat{W}_i \rangle$ and output channels $\hat{C}_i$.**

Stage i	Operator $\hat{\mathcal{F}}_i$	Resolution $\hat{H}_i \times \hat{W}_i$	#Channels $\hat{C}_i$	#Layers $\hat{L}_i$
1	Conv3x3	224×224	32	1
2	MBConv1, k3x3	112×112	16	1
3	MBConv6, k3x3	112×112	24	2
4	MBConv6, k5x5	56×56	40	2
5	MBConv6, k3x3	28×28	80	3
6	MBConv6, k5x5	14×14	112	3
7	MBConv6, k5x5	14×14	192	4
8	MBConv6, k3x3	7×7	320	1
9	Conv1x1 & Pooling & FC	7×7	1280	1