# Predicting culture and personality in online courses

Sylvio Rüdian
Humboldt University of Berlin
Weizenbaum Institute for the Networked Society
Berlin, Germany
ruediasy@hu-berlin.de

Gergana Vladowa
University of Potsdam
Weizenbaum Institute for the Networked Society
Potsdam, Germany
gvladova@lswi.de

Gunay Kazimzade
Technical University of Berlin
Weizenbaum Institute for the Networked Society
Berlin, Germany
gunay.kazimzade@tu-berlin.de

Jana Gundlach
University of Potsdam
Weizenbaum Institute for the Networked Society
Potsdam, Germany
janagundlach@uni-potsdam.de

Niels Pinkwart
Humboldt University of Berlin
Weizenbaum Institute for the Networked Society
Berlin, Germany
pinkwart@hu-berlin.de

## ABSTRACT

Online courses support learners to engage in distance learning. One emerging trend of the educational community is their personalization. Individual cultural characteristics and personality traits that influence individuals' behavior in online courses have not yet been examined in detail. It is often practically impossible to collect a lot of personal information regarding personality or culture in online courses. Therefore, it is necessary to fill in a comprehensive questionnaire. We show how accurately personality and cultural traits can be predicted by behavior in an online course. The paper reports exploratory data-informed work. We use a neural network with behavioral data as input. In case of successful prediction, instructors can use these items to define targeting groups as a pre step for personalization. Our results show, for example, that long-term orientation can be predicted best by an individual's behavior. It corresponds to the ability and attitude of the individual to focus on the future. Learners with high long-term orientation will spend longer periods of time in class preparing to successfully complete related exercises. We discuss our findings from an interdisciplinary perspective and propose perspectives for further research on personalization.

## Keywords

Personalization, online courses, e-learning, Big Five, personality, culture, machine learning

## 1. INTRODUCTION

Personalization in online courses is a trending topic of the educational community. If we consider lifelong learning, online courses have to provide support for a heterogeneous user base. The diversity of learners requires methods for adaption as there is no one-size-fits-all learning environment.

To be able to support learners with the knowledge and skills needed to succeed in a rapidly changing world, personalized online learning is one of the fast-growing research directions. Considering the massiveness of the online learning resources it is essential to investigate on the impact of culture and personality of the users in their experience of adapting online learning environments. Therefore, we need information about culture or personality traits, which are still missing in learning environments. Researchers focus on drop-out rates or final outcomes. This information can be predicted based on clickstream data of participants that was previously collected during an online course [1]. Using the data, instructors have the ability to help students at risk.

By observing the industry, websites have the opportunity to collect clickstream data as well. This can be used to predict demographic data, which allows them to separate users into groups of customers with similar attributes [2]. Marketers use this prediction to optimize the process for profit maximization. Online courses can be seen as a special category of websites with similar opportunities for optimization [3]. Instead of maximizing the profit, online courses follow the aim to teach, apply and test participants for knowledge transfer. Alternatively, the motivation using a learning environment could be optimized.

Classical educational recommender systems support users in finding learning material that could be beneficial for reaching their desired goals [4]. This is a macro view of personalization as it tries to find learning resources that the user potentially is looking for. By looking at the micro level of personalization we consider single online courses which could be optimized for individuals. People have different personalities, cultural background and learning styles. Thus we aim to suggest adding new predictable items to a user model that can further be used to personalize online courses at micro level [5].

Culture is a shared system of values [6] [7] [8]. The recent advancements in modernization have been identified as erasing cultural differences [9]. Increased globalization is anticipated to cause hybridization [9] [10]. Additionally, collecting data on cultural, ethical and national belonging may not always be possible or is morally questionable. However, in line with our aim to find out how online courses can be adjusted to fit the individual user's needs, information on national or cultural belonging is of great interest.

Learning has a strong connection with the culture of individuals and groups. Therefore, the educational systems of one country are not always applicable in another country which has different values, norms and standards [11] [12]. In order to predict culture in an online learning context, we approach the model developed by G. H.

Hofstede [13] which is identifying "cultural dimensions" which serve as measurement instruments of different cultures [14], and validated to CVScale which is further applied in this research. The cultural dimensions applied are Power Distance (PO), Uncertainty Avoidance (UN), Collectivism (CO), Long- Term Orientation (LT) and Masculinity/Femininity (M). As cultural traits do not often change in life, we have to consider these items for long-live learning. The tendencies of collectivism, uncertainty avoidance and the high power distance of Eastern cultures have been found in online learning environments [15] [16]. Bonk and Kim's study [17] shows the dominance of social interactions among Korean students at the outset of their online collaboration, which demonstrates their cultural inclination toward emphasizing relationships over tasks. Using these items as relative stable factors in life, personalization based on these characteristics can be an advantage because we can learn them by using the system without the necessity of a comprehensive questionnaire.

Instructional design community debates a lot about the impact of personality and culture in the personalized learning construction [18] [19] [20] [21]. Personality and culture of the learner has a strong correlation with the different learning styles, efficiency of the learner and motivation during the learning process [22]. Thus, predicting personality type and cultural characteristics of the learning can benefit to the customization of online courses with respect to the design and structure of the online learning materials.

The research objective of this paper is to explore the ability for prediction of culture / personality traits in online courses. Therefore, we examine: Which considered traits can be predicted by behavior in a linear online course?

The paper is structured as follows. The next section describes related work according to studies of learning and cultural / personality traits. Section 3 describes our methodology, followed by our results. In section 5 we discuss our results and explain our decisions made. Section 6 proposes some ideas for further investigations, followed by our conclusion.

## 2. RELATED WORK

In an experimental study, Makhija et al. [23] explored the links between demographic factors, personality, behavioral engagement and culture in relation to academic engagement. They used questionnaires to get dimensions of personality (Big Five Factor Model) and demographic factors. Cultural information was derived by asking participants which culture they belong in and was limited to ask for the country, where people currently live. Academic engagement was measured by using variables like received grades and time they spent on completing the tasks. Behavioral engagement was represented by students' attendance, participation in class and extracurricular activities.

Kloft et al. [1] used clickstream of an online course to predict dropouts. With considered scalar features they achieved an accuracy between 72% and 87%. This study shows that behavior can be used to predict dropouts. The resulting information about potential dropouts can help instructors to detect students that aim to drop out of the course.

Cultural background is an important concept with respect to the way of thinking, performing and learning of a specific group of people. Hence, investigating the cultural component in online learning and its connection to design patterns of the learning environment is crucial. Inclusivity of e-learning systems allows users across the world to access quality education. Thus, the relationship between users' cultural backgrounds and e-learning systems has been a topic of research of several researchers.

There is a strong connection between cultural dimensions and behavior during online learning. For the last two decades, researchers investigated several qualitative and quantitative analysis on the impact of cultural dimensions from G.H. Hofstede [13] to the learning and usability, behavior and outcomes of an online learning system.

With respect to the impact of culture to offline learning, research of Liu [24] discusses the intersection of the Hofstede Dimensions and the Cultural Dimensions within the context of the Learning Framework. The paper refers to Bonk et al. [17] which proposes that the power distance dimension alongside collectivism and uncertainty avoidance leads to the dominance of social interactions and an emphasis of relationships over tasks for Korean students. Additionally, Hofstede [25] refers to a potential heavy reliance on instructors and textbooks for people with a high power distance dimension.

Individualism has a strong connection to activeness in class to express themselves, to appreciate diverse opinions in learning, and to be self-motivated. Further, the masculinity dimension connects strongly with the high level of and desire for recognition. Furthermore, research that learners who avoid uncertainty are usually preferred receiving answers from structured learning activities.

McLoughlin [26] states that the flexibility of learners from mixed cultures in the e-learning systems is often limited. Most of those systems are adapted to the specific groups' need, learning style and their learning requirements. Another study from Downey et al. [27] focuses on the relationship between national culture and the usability of an e-learning system. They integrate Hofstede's cultural dimensions and Nielsen's usability attributes into the usability study of the e-learning materials and highlight the connection between each cultural dimension and its impact on usability.

During the study of [28] with Arab students who were examined during online learning, participants expressed their fear and anxiety of taking online courses because they equated online learning with independent learning which is capturing Arab culture's high uncertainty avoidance [29]. The study of [30] examining Jamaican and Canadian women's online learning experiences indicates the groups' cultural expectations regarding women's roles in the home and how it restricts their engagement and learning. Other studies [31] emphasized a strong uncertainty avoidance of Chinese students during online learning. They were constantly asking for "rules and instructions" and if there are any rituals for them to follow. With respect to usability and design, there have been a lot of studies regarding the impact cultural background of the user to the design preferences [32] and usability of the interfaces and online systems [33]. Research of Downey (2007) investigates how cultural dimensions are interconnected with the usability of e-learning systems. The study analyzes the cultural dimensions with respect to learnability, error rate, and user's satisfaction and exploring the relationship linked to power distance, individualism and collectivism, femininity/masculinity and uncertainty avoidance.

Cultural traits in online courses were investigated on its impact on communication difficulties [34]. Other studies focused on critical thinking, harmony, affection, compassion, emotionality, frustration, participation, success and performance [35]. According to Strang [36], culture is not cross-related to final grades. But grades can be predicted based on students' behavior [37]. Research on relations between culture and the behavior limit culture to the country where participants live [23]. This is a very general view concerning culture. Hofstede's cultural dimensions have not been used yet in online courses and have not been examined regarding personalization. The is a gap in research. We want to bridge the gap by showing that cultural dimensions by Hofstede [14] can be predicted due to behavioral patterns in online courses. We also want to compare the accuracy of predictable items with personality traits and demographic data.

In order to optimize individuals' learning processes, a lot of information about individual characteristics and their effects on learning and behavior is needed. An online course can only be individualized on the basis of certain realizable characteristics of the user. The single learner with all the unique complexity of his individuality cannot be captured. One way to describe and analyze a person is by personality traits. In personality psychology, the most frequently used taxonomies of personality traits are the Big Five personality dimensions. After decades of research, they were developed by consensus with the aim of enabling the investigation of specified areas of personality traits rather than examining many specific attributes that make people unique. The dimensions of the Big Five were developed based on natural language terms used by people describing themselves or others (for an extended overview of the development of the taxonomy view [38]). In addition, it can and will serve as a starting point for further research and theory development, explanation and revision of the taxonomy according to context [38]. The present framework of the Big Five is mainly the result of the work of Goldberg [39], McCrae and Costa [40].

Komarraju et al. [41] investigate the influence of personality on learning styles in the context of academic achievement. Conscientiousness and agreeableness were found to be positively related to all four learning styles (synthesis analysis, methodological study, factual fidelity and elaborative processing), while neuroticism was found to be negatively related to all four learning styles. Extraversion and openness are positively related to the elaborate processing. Furthermore, the relationship between openness and average grade is mediated by reflective learning styles (synthesis analysis and elaborative processing). Relevant studies on education and work performance support the five-factor model and its influence on several work-related constructs [42] [43] [44]. Extraversion, conscientiousness and openness are positively related to training proficiency (defined as training performance, productivity data and time to completion of training outcomes), whereby conscientiousness is explicitly associated with learning motivation and neuroticism - negatively with learning motivation (e.g. Colquitt and Simmering [45]; Colquitt et al. [46] [47]).

We decided to concentrate on only three of the five major personality traits, namely conscientiousness, openness and neuroticism. Most studies on the context of learning and personality show the strong impact of the three traits. The reasons for this decision are, on the one hand, the results of the studies presented - most of them point precisely to these factors as the most important influencing factors and as linked to learning behavior and output. On the other hand, we also take into account the particularities of our study - an online course that aims to examine a participant's learning process, acting separately and without any interaction with other participants or a teacher. In this context, the two personality traits extraversion and agreeableness were removed as not being relevant for our learning process.

However, there is another reason for abandoning the two dimensions: Our pretest has shown that participants find the Big Five questionnaire, consisting of 50 questions, too long, leading to breaks and useless results. Therefore, it was necessary to reduce the number of questions. We also considered using TIPI as a shorter version of the Big Five questionnaire as proposed by Makhija et al. [23]. However, this short questionnaire does not meet the requirements of our study for the following reasons: 1) validation or learning studies are still missing and 2) TIPI cannot provide a faceted picture of a single person, which allows the use of the longer questionnaire (Big Five) [48]. Specifically, Gosling et al. [49] point out that TIPI is "offered for situations where very short measures are required, personality is not a priority, or researchers can tolerate the somewhat diminished psychometric characteristics of very short measures", which is not consistent with our research objective. The Big Five personality model distinguishes five dimensions of personality (cf. Barrick and Mount [42]; Bidjerano and Dai [50]) and we consider the following:

- **Openness to experience (O):** Individuals are imaginative, curious, flexible, creative, seeking novelty, original. With regard to learning contexts, it was found that openness is linked to a deep approach to learning, elaborative learning [51], [52], meaning-directed learning, and constructive learning [53].

- **Conscientiousness (C):** The individual is systematic, efficient, organized, reliable, responsible, diligent, persistent, self-disciplined. In the learning context it is associated with motivation, effort and perseverance [54] as well as with methodological and analytical learning [51].

- **Neuroticism (N):** Individuals are disturbing, anxious, insecure, depressed, self-conscious, moody, emotional, unstable. Neuroticism is associated with poor critical thinking skills, analytical ability and conceptual understanding. Individuals with high neuroticism probably have a superficial approach to learning - to focus on memorizing and superficial features of the material being studied, rather than gaining a deeper, meaningful understanding of it [55].

We aim to use clickstream data and examine the prediction of personality or cultural traits. In comparison with the prediction of dropouts, we can detect behavioral features that cause them. Our experimental study shows that clickstream data can be used for prediction of our items and thus targeting groups can be detected by specific behavioral patterns. Exploring behavioral patterns can help instructors to personalize different areas of an online course, based on targeting traits.

# 3. METHODOLOGY

Our approach focuses on finding features that can be predicted by behavioral data. We assume that features that have a known influence on learning behavior can be predicted in an online course. These features can be used for further studies to detect learners needs according to personality or cultural traits in an online course, which could be different for various targeting groups. Thus our considered items can be used for personalization under condition that the targeting groups' learners' needs are known. We use machine learning to predict our items based on the behavior. This shows whether our items can be predicted and how well they perform in a real-world scenario. The resulting list contains each item and the corresponding accuracy that could be achieved. Sorting by accuracy gives us an idea which items are predictable due to behavior. Items where the machine learning algorithm has a bad accuracy still require completing a questionnaire if we need the traits.

First, we give an overview of our online course. We used a commercial online course to conduct our study. It consists of tree sub lectures ($L_1$, $L_2$ and $L_3$) that include information pages [P] and interactive tasks (multiple-choice question [$T_{MC}$], finding the right sequence [$T_{SE}$], fill in blanks [$T_{BL}$], open task [$T_{OT}$]), followed by a questionnaire [Q]. It has the following structure:

$L_1 = \{P_{Basics}, P_{Basics}, P_{Example}, P_{Example}, T_{MC}, T_{SE}, Q_{Culture}\}$
$L_2 = \{P_{Basics}, P_{Example}, P_{Example}, P_{Basics}, P_{Basics}, \{T_{MC}, T_{MC}, T_{BL}, T_{MC}\}, Q_{Personality}\}$
$L_3 = \{P_{Basics}, P_{Basics}, P_{Basics}, P_{Example}, T_{SE}, T_{OT}, Q_{Feedback}\}$

We used Moodle as technical learning platform and structured all contents and tasks. Interactive tasks were implemented with the plugin H5P[1]. This plugin contains different methods of tasks with the ability to give interactive feedback.

All questionnaires (for culture, personality and feedback) were placed at the end of each lecture. We aimed to acquire some participants that are interested in the online course's topic itself and not having financial interests. By having any of the questionnaires at the beginning, the dropout rate would be much higher. Thus, we decided to place them at the end.

The participant's behavior was captured by their interactions with the online course. From every page view we logged the time until the user clicks on another page or task. We also captured how often the user viewed pages to detect multiple views. For all tasks we logged the time to finish and we logged the success rate of the answers. The task $\{T_{MC}, T_{MC}, T_{BL}, T_{MC}\}$ of $L_2$ is a collection of tasks, where we could extract the overall time only with the success rates for each containing task individually. Texts entered in the open task are captured additionally to extract their lengths. We also logged the length of the feedback because we assumed that this information could have an impact to our items. We define all this data as our behavioral data B.

On the other hand, we used the answers of the questionnaires to apply cultural and personality dimensions because we wanted to identify influences of these dimensions to the behavior. Apart from the culture and personality, we also collected demographic information (age, gender). As we detected the time that all participants need to view single pages, we also logged the browser header to split our data into two datasets (mobile device and desktop). This split is necessary due to different screen sizes, which may lead to different reading time because of the necessity to scroll down on small screens. We call this data D. The resulting dataset was mapped into the vector B, consisting of 13 items for page view durations, 13 items for repeated page views, 5 items for task durations and 8 items for the task success rates. D consists of 5 cultural dimensions, 3 personality dimensions, 2 demographic information. The 5 cultural dimensions were calculated according to CVScale [8], our considered three personality traits were calculated with given formulas of the Big Five test [38].

Various other scales such as Schwartz/Rockeach; GLOBE[2] (Global Leaderhip and Organizational Behavior Effectiveness), The World Value Survey [3]and a scale by Minkov [56] as well as various adaptations are in use to quantify cultural values. Among these variables such as GLOBE and the CVScale are built on the core of Hofstede's dimensions. The CVScale comprises a "26-item five-dimensional scale of individual cultural values" [8] that estimates the Hofstede cultural values at the individual level. Being regularly used [57] [58], it shows reliability, validity and generalizability across samples and nations [8]. Also, it applies to a broader context beyond management [8]. It has been mainly criticized for using the same labels as within the Hofstede model, describing differing concepts [15]. However, the pool of items used for the respective scale was adapted and build upon modified items from the HERMES values questions, which are Hofstede's original questions [29], the Values Survey Module 1994 [14] and additional Hofstede works [8]. Some additional items from other construct scales were used where applicable and items were refined until the scale was valid and reliable [8].

Next, we designed a neural network for each item and optimized its hyperparameters to achieve the best accuracy. Grid Search [59] helped us to find the best hyperparameters automatically. We transferred all considered traits to three classes, because the general idea of Big Five and CVScale is not to get exact values to describe personality or culture. Instead, all values are used to classify people, e.g. the Big Five is used to understand social traits of employees (range: 1-40). Thus, we defined three classes: low (1-13) – medium (13.1-26) – high (26.1-40) for personality traits and low (1-2.33) – medium (2.331-3.66) – high (3.661-5) for cultural traits (range: 1-5). Figure 1 shows an example of the derivation of openness to experience, based on three classes as defined before. All other distributions look the same and are nearly equal distributed, which avoids overfitting.

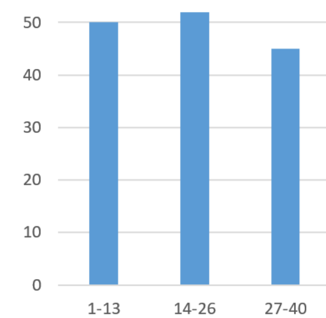


**Figure 1. Distribution of openness to experience.**

We used the 5-fold-cross-validation (5f-CV) that splits our data into five parts, and we built the model with four of them and tested with one part. Thus, our resulting accuracy is the result of predicting on previously unseen data. We rotate the test part and average the final accuracy to get an appropriate generalizable value.

Instructors can predict these items in order to define various targeting groups that share similar learning styles due to similar cultural backgrounds or personality traits. According to the learners needs these groups can use different versions of an online course, which might help to achieve better learning goals. This adjustment can be providing different contents or usability changes in order to optimize the online course for targeted personalization. What the concrete design decisions between multiple versions should be, has to be examined in further investigations.

# 4. RESULTS

The online course itself was about a technical related topic (Search Engine Optimization[4]). We distributed the free online course in different social media's groups related to business, marketing and startups. Additionally, we distributed the course via several university's mailing lists. Finally, 142 participants took part in our study.

We limited our study to desktop users to eliminate potential time biases due to different screen sizes. By examining data, less than 1% used the mobile version of our online course.

[1] https://h5p.org/

[2] https://globeproject.com

[3] http://www.worldvaluessurvey.org/wvs.jsp

[4] https://course.seorld.com

**Table 1. Accuracy for each item**

| Class | Item | Configuration of Neural Network | | 5f-CV in % |
|---|---|---|---|---|
| | | **Nodes** | **Activation** | |
| Culture | PO | 40<br>5<br>3 | softmax<br>softmax<br>hard_sigmoid | 82.22 |
| | UN | 40<br>5<br>3 | hard_sigmoid<br>softplus<br>hard_sigmoid | 84.44 |
| | ID | 40<br>5<br>3 | hard_sigmoid<br>softplus<br>hard_sigmoid | 88.89 |
| | **LT** | 40<br>5<br>3 | sigmoid<br>relu<br>hard_sigmoid | **93.33** |
| | MA | 40<br>5<br>3 | sigmoid<br>relu<br>hard_sigmoid | 86.67 |
| Personality | **O** | 40<br>50<br>3 | sigmoid<br>softmax<br>softplus | **91.67** |
| | C | 40<br>50<br>3 | hard_sigmoid<br>softplus<br>hard_sigmoid | 83.33 |
| | N | 40<br>50<br>3 | sigmoid<br>softplus<br>sigmoid | 88.33 |
| Demographic data | Age | 40<br>5<br>5 | hard_sigmoid<br>softplus<br>hard_sigmoid | 85.33 |
| | Gender | 40<br>5<br>3 | hard_sigmoid<br>sigmoid<br>sigmoid | 83.33 |

To predict every item, which is comparable to a real-world scenario, we created a neural network for each, consisting of a three-layer architecture (epochs: 500, batch_size: 250). The Grid Search approach optimized the hyperparameters for us, shown in Table 1. Thus, we could find the optimal accuracy in 5-fold-cross-validation (5f-CV), that could be achieved if being applied in an application.

We assumed that some features might not be predictable as there are no detectable behavior patterns. Our results in Table 1 show that all accuracies for prediction are at least 82%. This is a surprising result which shows that all our items can be predicted by behavioral data. The resulting accuracies are comparable to the accuracy to predict dropouts in online courses. This topic has been investigated a lot by the educational communities and authors are able to achieve accuracies between 72% and 87% [1]. Most of our accuracies are even better for our considered items.

The cultural long-term-orientation index and the openness to experience have the highest accuracy. These items' classes can be accurately predicted for most participants.

Online courses mostly have access to demographic data only. Accuracy is not bad but there are other items that can be predicted better. Since our research question was to identify a subset of items, we can conclude that all items can be predicted based on behavior.

Depending on the participants' acceptance of a maximum number of questions, we can choose the best predictable subset of items with corresponding questions. If we use the best three items in a real-world scenario, this selection requires a questionnaire with 22 questions (6 for long-term orientation index, 10 for openness to experience and 6 for individualism). In contrast, to get the characteristics of all 8 items, answers of 56 questions are required (26 for culture and 30 for three personality traits) or 76 questions if we use the complete Big Five questionnaire plus CVscale. Questionnaires with less questions might be used, but they have to be evaluated first.

# 5. DISCUSSION

Cultural dimensions in origin were identified with respect to their influence on human interactions in established systems, social organizations, and education. These were the factors that have an impact on the usability of online learning systems, however, we have to state that these cultural variables as defined by Hofstede [13] were not designed specifically for studying usability and behavior in online learning. In the study of Zaharias et al. [60] researchers analyzed a connection between collectivism and learnability of a web-based testing system. Another study by Downey [27] investigates the usability attributes as learnability, efficiency, memorability, errors, and satisfaction and the results show that participants from collectivist cultures showed strong, statistically significant levels of satisfaction with the system they used. These participants' results had strong correlations between their low uncertainty avoidance score and their higher errant click rates.

Also, individuals from cultures with high power distance indicator scores usually made more erratic mouse clicks while using the system. However, it is important to mention that not all indicated studies have a focus on online learning and each of them has different research methodologies and findings as well as different participants.

The experimental study was limited by the number of participants that finished the online course and the questionnaires both. The results become more accurate, the more participants take place. Thus, we aim to continue our study with more participants. Over 99% of the 142 participants did not used a Smartphone to take part. From the practical perspective, the experiments should be applied with mobile users, those results might be different from desktop users. This can help instructors to adjust contents by splitting the targeting groups by the used device as well.

According to Hu et al. [61], gender and age can be predicted on general websites as well. At websites as an unstructured environment, they achieved an accuracy of 79.7% on gender and 60.3% on age. Our online course has a linear structure with non-sparse data, which makes it easier to predict gender and age. Thus, our accuracy is better. Our result shows that personality and cultural traits can be predicted even better, limited to our study by using behavioral data of the online course. In our study, we could benefit from the linear structure. If behavioral data becomes more unstructured due to applying educational recommender systems, our prediction rates will become worse.

Although, the indulgence versus restraint measure by Hofstede was not included in our questionnaire, one could assume that there might be a strong link to online learning. Indulgence is concerned with any behavior that fosters fun and allows for the pursuit of desires and enjoyment, whereas restraint indicates ones pulling oneself together in order to comply with social norms [14]. This dimension could have contributed to the click through patterns of the online course. The course displays different aspects, such as gamification concerning the tasks as well as survey submission and text reading which may potentially be rather tedious and hence requiring some degree of discipline or restraint. The used questionnaire, the CVScale was not designed to get the indulgence index. Further research should take this dimension into account by using further questionnaires.

Concerning personalization, we have to understand the relation between our considered traits and learning goals. According to Hofstede, long-term orientation is a time based perspective and knowing this dimension for every participant can help to understand how they perform in an online course. Being able to predict long-term orientation gives instructors information about culture, which is linked to different variants of learning. Students that have a high openness to experience can use experience-based learning and might perform better, while others need more structured knowledge to achieve the same level of knowledge.

If we cluster the behavioral data according to our considered traits, instructors can detect differences for various targeting groups. We examined the trait openness to experience (O) and could see that the average time spent on one specific page is the following: 111 sec. (low O), 156 sec. (medium O) and 142 sec. (high O). Participants with low OP spend less time on the page, thus the instructor could optimize the content for this specific targeting group. For personalization it is important that claims like this will be triangulated with achievement data to optimize the online course concerning the learning goal. How the optimization itself can take place is the investigation of further research.

Our experiment was limited to one specific online course. To predict our considered items in another online course, we still require a training step. This is the general training problem that prediction tasks like predicting dropouts or final outcomes have in common. To generalize our approach for a wider usage without the necessity of a training step, general behavior patterns must be found that have an importance in prediction of our traits. Therefore, we have to repeat our experiment with other online courses that have a different structure to find behavioral similarities for prediction.

# 6. FUTURE WORK

Our future explorative research in the context of personality and learning includes the application and testing of the method with other personality tests. Two tests have already been identified as relevant: 1) The Myers-Briggs indicator [62]: This test consists of 94 items developed on the basis of the four bipolar discontinuous scales of the theory of Carl Jung [63]: Introversion-Extraversion, Sensations-Intuition, Thinking-Feeling and Judging-Perceiving. The classification of respondents into one of the 16 personality types is based on the highest score obtained for each bipolar scale. 2) The Keirsey Temperament Sorter [64] developed 16 personality types based on works by Socrates and Plato (with their four temperament models - Artisan (iconic), Guardian (pistish), Idealist (poetic) and Rational (diatonic)). He has divided the four temperaments into two categories (roles), each containing two types (role variants). We could examine whether these traits can be predicted by behavior as well.

We can use existing studies on the correlation between the Big Five and these two personality tests and on the correlation between these personalities and learning styles. We also want to test the Big Five's two characteristics - agreeableness and extraversion - in an appropriate collaborative learning environment.

Additionally, given that the order of the CVScale was adapted to avoid order response bias, additional scale validation could increase its reliability. One could even consider combining the validation of the two scales of personality and national cultural dimensions in order to rule out any correlations which were previously pointed out by Hofstede [14]. With respect to cultural dimensions, further research could include other ways of data collection as well as the combination of relevant cultural background specifications and questionnaires. To further extend our research we also aim to analyze behavior within online learning from the perspective of the users with special needs and disabilities. Furthermore, the research could be further expanded with the additional parameters as emotional engagement, and cultural specifications combining the subcultures and personality restrictions of the learners.

To understand how many questions can be used in a real-world scenario, we need an additional study to examine the acceptance rate concerning the amount of questions we ask. If we have a concrete number of accepted questions, we are able to give a recommendation how many traits could be considered for personalization.

# 7. CONCLUSION

In this paper we presented an experimental study to explore the prediction of culture and personality traits based on the behavior within online courses. We used an online course with additional questionnaires to get necessary data of our considered characteristics. We trained neural networks to show how all dimensions can be predicted in a real-world scenario. We followed the idea that, if our items could be predicted by the behavior. Unlike assumed, there is no item that cannot be predicted and thus no item can be ignored in general. Two items could be predicted best (long-term orientation and openness to experience). The cultural item “power distance” has the worst accuracy. This validates our assumption that this item can be predicted word by behavior in an online course. We conclude that instructors could focus on the best two items for prediction and further usage in online courses.

Our study does not show how the online course should be adjusted. This has to be examined in further studies, but knowing which traits are predictable can help instructors to split users into different targeting groups, which are an important base to personalize online courses. Thus, our approach helps to support lifelong learning with personalized online courses for a wide range of people with different personalities and cultural backgrounds.

Previous research of predictions in online courses still ignored cultural dimensions. Our experiment has shown that culture can also be considered at an individual level, instead of using the country only, where participants currently live. We gave reasons for the decisions we made for our experimental study and discuss the relation of culture and personality with respect to learning in an online course. Cultural and personality traits should be the focus of further studies of personalized learning in online courses.

# 8. ACKNOWLEDGMENTS

This work was supported by the German Federal Ministry of Education and Research (BMBF), grant number 16DII116 (Weizenbaum-Institute). The responsibility for the content of this publication remains with the authors. We would like to thank the company seorld for providing access to their online courses.

# REFERENCES


[1] M. Kloft, F. Stiehler, Z. Zheng and N. Pinkwart, "Predicting MOOC Dropout over Weeks Using Machine Learning Methods," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNL*, Qatar, Association for Computational Linguistic, 2014, pp. 60-65.

[2] A. L. Montgomery, S. Li, K. Srinivasan and J. C. Liechty, "Modeling Online Browsing and Path Analysis Using Clickstream Data," in *Marketing Science 23(4)*, 2004, pp. 469-631.

[3] A. Carr-Chellman and P. Duchastel, "The ideal online course," in *British Journal of Educational Technology 31(3)*, 2002.

[4] O. C. Santos, "Educational Recommender Systems and Technologies: Practices and Challenges," 2011.

[5] H. Drachsler, H. Hummel and R. Koper, "Personal recommender systems for learners in lifelong learning: requirements, techniques and model," in *International Journal of Learning Technology, 3(4)*, 2008, pp. 404-423.

[6] R. Fischer, "Where is culture in cross-cultural research? An outline of a multilevel research process for measuring culture as a shared meaning system," in *International Journal of Cross Cultural Management, Vol. 9 No. 1*, 2009, pp. 25-49.

[7] S. Schwartz, "A theory of cultural value orientations: explication and applications," in *Comparative Sociology, Vol. 5 Nos 2-3*, 2006, pp. 138-182.

[8] B. Yoo, N. Donthu and T. Lenartowicz, "Measuring Hofstede's five dimensions of cultural values at the individual level: Development and validation of CVSCALE," in *Journal of International Consumer Marketing 23(3-4)*, 2011, pp. 193-210.

[9] J. N. Pieterse, "Globalisation and culture: Three paradigms," in *Economic and political weekly*, 1996, pp. 1389-1393.

[10] H. J. Hermans and H. J. Kempen, "Moving cultures: The perilous problems of cultural dichotomies in a globalizing society," in *American psychologist53(10)*, 1998, p. 1111.

[11] Hall and Hall, "Understanding Cultural Differences," 1990.

[12] B. Collis, "Designing for Differences: Cultural Issues in the Design of WWW-Based Course-Support Sites," 2002.

[13] G. Hofstede, "Culture's Consequences: International Differences in Work-Related Values," Beverly Hills, CA: Sage., 1980.

[14] G. Hofstede, "Dimensionalizing cultures: The Hofstede model in context," in *Online readings in psychology and culture2(1): 8*, 2011, p. 9.

[15] H. Ku and L. L. Lohr, "A case study of Chinese students' attitude toward their first online learning experience," in *Education Technology Research and Development, 51(3)*, 2003, pp. 94-102.

[16] M. Wang, "Designing online courses that effectively engage learners from diverse cultural backgrounds," in *British Journal of Educational Technology, 38(2)*, 2007, p. 294–311.

[17] K. Kim and J. B. C., "Cross cultural comparisons of online collaboration," in *Journal of Computer-Mediated Communication 8(1)*, 2002.

[18] D. &. O. E. Song, "Learning styles based on the different cultural background of KFL learners in online learning," in *Multimedia-Assisted Language Learning, 14(3)*, 2011, pp. 133-154.

[19] L. Mestre, "Accommodating Diverse Learning Styles in an Online Environment," in *Reference & User Services Quarterly 64(2)*, JSTOR, 2006, pp. 27-32.

[20] D. R. Annette Vincent, "Personalize training: determine learning styles, personality types and multiple intelligences online," in *The Learning Organization, 8(1)*, MCB UP Ltd, 2001, pp. 36-43.

[21] V. Eyharabide, I. Gasparini, S. Schiaffino, M. Pimenta and A. Amandi, "Personalized e-Learning Environments: Considering Students' Contexts," in *Education and Technology for a Better World (WCCE)*, Springer, 2009.

[22] P. Mohammed and P. Mohan, "Dynamic cultural contextualisation of educational content in intelligent learning environments using ICON," in *International Journal of Artificial Intelligence in Education*, Springer, 2014, pp. 249-250.

[23] A. Makhija, D. Richards, J. d. Haan, F. Dignum and M. J. Jacobson, "The Influence of Gender, Personality, Cognitive and Affective Student Engagement on Academic Engagement in Educational Virtual Worlds," in *AIED*, London, Springer International Publishing AG, 2018, pp. 297-310.

[24] X. Liu, S. Liu, S. Lee and R. Magjuka, "Cultural Differences in Online Learning: International Student Perceptions," in *Educational Technology & Society, 13*, 2010, pp. 177-188.

[25] G. Hofstede, "Cultural differences in teaching and learning," in *International journal of Intercultural Relations 10*, 1986, pp. 301-320.

[26] C. McLoughlin, "Culturally responsive technology use: developing an online community of learners".

[27] S. Downey, R. Wentling, T. Wentling and A. Wadsworth, "The Relationship Between National Culture and the Usability of an E-Learning System," 2004.

[28] A. Al-Harthi, "Distance Higher Education Experiences of Arab," in *International Review of Research in Open and Distance Learning Journal, 6 (3)*, 2005, pp. 1-14.

[29] G. Hofstede, "Cultures and Organizations: Software of the Mind," London: McGraw-Hill, 1991.

[30] P. A. Gouthro, "Assessing Power Issues in Canadian and Jamaican Women's Experiences in Learning via Distance in Higher Education," in *Teaching in Higher Education 9(4)*, 2004, pp. 449-461.

[31] S. N. Smith and P. J. Smith, "Implications for Distance Education in the Study Approaches of Different Chinese National Groups," in *International Journal of E-Learning and Distance Education*, 2000, pp. 71-84.

[32] V. Evers, A. Kukulska-Hulme and A. Jones, "Cross-Cultural Understanding of Interface Design: A Cross-Cultural Analysis of Icon Recognition".

[33] J.-S. Lee, "Interactivity: A New Approach," in *Association for Education in Journalism and Mass Communication*, Phoenix, AZ, 2000.

[34] C. N. Gunawardena, P. L. Wilson and A. C. Nolla, "Culture and Online Education," in *Handbook of Distance Education*, 2003, pp. 753-775.

[35] S. Uzuner, "Questions of Culture in Distance Learning: A Research Review," in *International Review of Research in Open and Distance Learning 10(3)*, USA, 2009.

[36] K. D. Strang, "Beyond engagement analytics: which online mixed-data," in *Educ Inf Technol*, New York, Springer Science+Business Media, 2017, pp. 917-937.

[37] B. Xu and D. Yang, "Motivation classification and grade prediction for MOOCs learners," in *Journal Computational Intelligence and Neuroscience*, 2016.

[38] O. P. John and S. Srivastava, "The Big Five trait taxonomy: History, measurement, and theoretical perspectives. Handbook of personality: Theory and research 2," 1999, pp. 102-138.

[39] L. R. Goldberg, "The development of markers for the Big-Five factor structure.," in *Psychological Assessment, 4*, 1992, p. 26–42.

[40] R. R. McCrae and P. T. J. Costa, "Updating Norman's "adequacy taxonomy": Intelligence and personality dimensions in natural language and in questionnaires," in *Journal of Personality and Social Psychology, 49*, 1985, p. 710 –721.

[41] M. Komarraju, S. J. Karau, R. R. Schmeck and A. Avdic, "The Big Five personality traits, learning styles, and academic achievement," in *Personality and individual differences, 51(4)*, 2011, pp. 472-477.

[42] M. Barrick and M. Mount, "The Big Five personality dimensions and job performance: A meta-analysis.," in *Personnel Psychology, 44*, 1991, p. 1–26.

[43] M. Mount and M. Barrick, "Five reasons why the "big five" article has been frequently cited," in *Personnel Psychology, 51*, 1998, p. 849–857.

[44] A. Vinchur, S. J., F. Switzer and P. Roth, "A meta-analytic review of predictors of job performance for salespeople," in *Journal of Applied Psychology, 83*, 1998, p. 586–597.

[45] J. Colquitt and M. Simmering, "Conscientiousness, goal orientation, and motivation to learn during the learning process: A longitudinal study," in *Journal of Applied Psychology, 83*, 1998, p. 654–665.

[46] J. Colquitt, J. LePine and R. Noe, "Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research," in *Journal of Applied Psychology*, *85*, 2000, p. 678–707.

[47] D. A. Major, J. E. Turner and T. D. Fletcher, "Linking proactive personality and the Big Five to motivation to learn and development activity.," in *Journal of applied psychology 91(4)*, 2006, p. 927.

[48] J. J. Denissen, R. Geenen, M. Selfhout and M. A. Van Aken, "Single item Big Five ratings in a social network design," in *European Association of Personality Psychology, 22(1)*, European Journal of Personality, 2008, pp. 37-54.

[49] S. D. Gosling, P. J. Rentfrow and W. B. Swann Jr, "A very brief measure of the Big-Five personality domains," in *Journal of Research in personality, 37(6)*, 2003, pp. 504-528.

[50] T. Bidjerano and D. Y. Dai, "The relationship between the big-five model of personality and self-regulated learning strategies," in *Learning and individual differences*, *17(1)*, 2007, pp. 69-81.

[51] E. Geisler-Brenstein, R. R. Schmeck and J. Hetherington, "An individual difference perspective on student diversity," in *Higher Education*, *31*, 1996, p. 73−96.

[52] A. Slaats, J. Van der Sanden and J. Lodewijks, "Relating personality characteristics and learning style factors to grades in vocational education," in *Annual Meeting of the American Educational Research Association*, Chicago, 1997.

[53] V. V. Busato, F. J. Prins, J. J. Elshout and C. Hamaker, " The relations between learning styles, the Big Five personality traits and achievement motivation in higher education," in *Personality and Individual Differences, 26*, 1999, p. 129−140..

[54] T. Chamorro-Premuzic and A. Furnham, "Personality predicts academic performance: Evidence from two longitudinal university samples," in *Journal of research in personality*, *37(4)*, 2003, pp. 319-338.

[55] N. Entwistle, "Motivational factors in students' approaches to learning," in *R. R. Schmeck (Ed.), Learning strategies and learning styles*, New York, Plenum Press, 1988, p. 21−49.

[56] M. Minkov, "Cultural Differences in a globalizing world," Bingley, UK, Emerald, 2011.

[57] S. Kwok and M. Uncles, "Sales promotion effectiveness: the impact of consumer differences at an ethnic-group level," in *Journal of Product & Brand Management 14(3)*, 2005, pp. 170-186.

[58] K. J. Wilby, M. J. B. Z. Govaerts and D. H. J. M. Dolmans, "Exploring the influence of cultural orientations on assessment of communication behaviours during patient-practitioner interactions," in *BMC medical education 17(1)*, 2017, p. 61.

[59] "Tuning the hyper-parameters of an estimator," scikit-learn developers, 2017. [Online]. Available: http://scikit-learn.org/stable/modules/grid_search.html. [Accessed 30 07 2018].

[60] Z. P, V. K and P. K, "Designing online learning courses: implications for usability," 01 01 2001. [Online]. Available: http://www.japit.org/zaharias_etal022.pdf. [Accessed 02 02 2019].

[61] J. Hu, H.-J. Zeng, H. Li, C. Niu and Z. Chen, "Demographic Prediction Based on User's Browsing Behavior," in *Predictive Modeling of Web Users*, Canada, 2007, pp. 151-160.

[62] I. B. Myers, "The Myers-Briggs Type Indicator: Manual," Palo Alto, CA, Consulting Psychologists Press, 1962.

[63] C. G. Jung, "Psychological types (HG Baynes, trans., revised by RFC Hull)," in *The collected works of CG Jung*, *6*, 1971, pp. 510-523.

[64] D. Keirsey and M. M. Bates, "Please understand me: Character & temperament types," Del Mar, CA, Prometheus Nemesis Book Company, 1984, pp. 3-4.