Predicting Creativity in Online Courses

Conference Paper · July 2022

DOI: 10.1109/ICALT55010.2022.00056

3 authors, including:

Leo Sylvio Rüdian
Humboldt-Universität zu Berlin

Jennifer Haase
Universität Potsdam

Predicting Creativity in Online Courses

Sylvio Rüdian
Humboldt-Universität zu Berlin,
Department of Computer Science
Weizenbaum Institute
Berlin, Germany
orcid.org/0000-0003-3943-4802

Jennifer Haase
University of Potsdam
Chair of Business Informatics
Weizenbaum Institute
Potsdam, Germany
orcid.org/0000-0001-8450-7252

Niels Pinkwart
Humboldt-Universität zu Berlin,
Department of Computer Science
Weizenbaum Institute
Berlin, Germany
orcid.org/0000-0001-7076-9737

2022 International Conference on Advanced Learning Technologies (ICALT) | 978-1-6654-9519-6/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICALT55010.2022.00056

***Abstract*— Many prediction tasks can be done based on users' trace data. This paper explores divergent and convergent thinking as person-related attributes and predicts them based on features gathered in an online course. We use the logfile data of a short Moodle course, combined with an image test (IMT), the Alternate Uses Task (AUT), the Remote Associates Test (RAT), and creative self-efficacy (CSE). Our results show that originality and elaboration metrics can be predicted with an accuracy of ~.7 in cross-validation, whereby predicting fluency and RAT scores perform worst. CSE items can be predicted with an accuracy of ~.45. The best performing model is a Random Forest Tree, where the features were reduced using a Linear Discriminant Analysis in advance. The promising results can help to adjust online courses to the learners' needs based on their creative performances.**

Keywords*— *Prediction, Online Course, Trace data, Creativity

I. INTRODUCTION

Online courses are used as a flexible and efficient way for self-regulated learning. In particular, learners are diverse, and there is no one-size-fits-all environment. Providing a customized learning platform according to individual needs allows users with different learning preferences and various personal attributes to learn efficiently [1]. Personalization is not limited to knowledge and skills. The way of how to teach could be adapted based on person-related characteristics. One of these attributes has not often been studied in the digital learning context: creativity. To adapt learning environments based on individual competencies and preferences, the individuals' creative abilities must be assessed first to derive a user model. However, practical applications have shown that learners do not want to participate in person-related tests before starting an online course. Questionnaires include far too many items; they are too personal without any relation to learning or tasks require too much time without focusing on the scope of the course. Especially when assessing creative abilities, a test-like measurement approach hampers the flow of ideas being assessed. This is why an indirect way of using trace data to deduce individual creative abilities is of high interest to avoid disturbing the learner during the learning process and access creative features as a base for adaptions and recommendations. Digital games have already been used to measure individual creative abilities. Although such approaches are promising, they are still in their infancy [2]. Online courses provide the feature that data can be collected on how individuals interact with the learning material automatically, without additional conscious effort for the learner and teacher. It would thus be wise to make use of this information.

II. RELATED WORK

Classical learning management systems store all user interactions in logfiles. Dietz-Uhler and Hurn [3] distinguish two types of data: trace data, which includes clicks on learning material, and performance data, e.g., resulting in scores from interactive tasks. Meta-data can be derived as an abstract representation of user behavior based on data. This includes how much time learners spent with the learning material, whether they accessed them multiple times, or how well tasks of different difficulty levels were solved. Interactive online courses can be created without substantial technical efforts. Nowadays, it is unnecessary to program interactive tasks as the base for many didactical self-learning methods is already implemented by H5P. There, pre-defined templates can be filled with learning content. When learners interact with the tasks, performance data is gathered on item level. For each question, e.g., in a question set, the given answer can be accessed using logfiles. This offers a vast amount of data.

Examining related studies, trace, and performance data is often used for online course predictions. Kloft et al. [4] have shown that the dropout rate of participants can be predicted with acceptable accuracy. Performance data can be used to predict future success [5]. Other studies have shown that personality-based attributes are predictable from digital records [6]. Various studies in educational data mining have successfully demonstrated the use of trace and performance data as a valuable base for predicting success and motivational features [7]. Online courses have the advantage that many consistent features can be collected for single courses – as all participants of a concrete course will typically be guided through it. Thus, mainly a linear learning path exists, and many features from the course interaction can be used to train a predictive model. However, this linear learning path is just an optimum for some learners,

978-1-6654-9519-6/22/$31.00 ©2022 IEEE
DOI 10.1109/ICALT55010.2022.00056

not everyone. Adaptive courses could be a solution that differentiate the teaching method [8]. Creativity can be a factor that splits learners into groups, and they get the differentiated learning paths that best suit their abilities, motivational level, and thinking styles.

Creativity is a complex individual trait characterized by cognitive elements, personality traits, and social embeddedness [9]. The cognitive basis of creative abilities relies on two types of thinking: associative/divergent thinking (DT) vs. convergent thinking (CT). DT describes the mental ability to develop many ideas for an open-ended problem. In contrast, CT requires critical thinking to develop the one unusual solution to a closed problem [10]. The creative process commonly requires DT for generating novel ideas and CT for the reflection and evaluation of those ideas. People tend to prefer and perform differently on those measures [11] and improve their performances differently [12]. In the context of adapting learning materials, this comes with two implications: creativity can be adapted based on the skill level, prior experience, and the learner's preferred creative thinking style. Individuals like to solve creative tasks with free associations or rational considerations, which corresponds to DT and CT [13]. Performance-based adaptations can better match a preferred thinking style as performance is positively associated with it [14].

If a student feels competent and in line with the preferred thinking style, the motivation to approach a creative task increases [15]. Although task motivation is of direct situational nature and bound to the specific task at hand, a person's predisposition to engage in creative tasks is based on prior experiences with comparable tasks. This is addressed through an individual's *creative self-efficacy* (CSE), the belief in one self's creative abilities [16]. People are much more likely to engage in tasks if they assume they will master them and thus, expect a positive outcome. Self-efficacy beliefs are best improved through experience and feedback. Through creative performance, individuals can directly experience how they are (more or less) capable of performing the task. If the task is slightly demanding, a positive accomplishment will enhance individual beliefs of mastering them [17].

Adapting online courses based on creative abilities holds the potential to be applied to learning courses beyond a creativity scope [18]. As creative thinking (DT and CT) is a generic skill involved in tasks like math – as an example [19] – adaptions that are in line with thinking preferences can also be beneficial for learner's motivation in other learning contexts [20]. Learning material could be adjusted to match the creative ability level needed for learning tasks, and the individual preference for more open vs. closed problems. If tasks are mildly challenging, their accomplishment will positively impact self-perception. Too complicated, and the motivation to even try will be low. Too easy, and a positive achievement cannot boost self-perception (Yerk-Dodson-Law [21]). Thus, knowing students' (non)expertise levels and adopting learning material can positively impact their learning outcomes.

To our best knowledge, no research uses trace and performance data to predict individual creative skills and self-beliefs within an online course. Based on these findings, we examine the prediction of creative performances and self-beliefs using the learner's clicking behavior in an online course. Therefore, we focus on the research question, which creative-related aspects, that are represented by different metrics, can be predicted with an appropriate accuracy?

III. Methodology

A higher education online course teaching the basics of the creativity concept was developed based on Moodle. We had full control and access to the trace data collected within the learning management system. Multiple modules, including pages, quizzes, and interactive tasks using H5P were part of the course. Questions about the course content were given, followed by content-related interactive tasks. They contained several multiple-choice questions about the prior presented contents and two gamified tasks, namely a "memory" game with pairs of terms and their definition and a sequence game, where innovations must be brought into chronological order. Throughout the course, several creativity assessments were administered. Based on a pre-test, these measurements were not recognized as such by the course participants but were perceived as part of the interactive tasks, which should increase the task performance as no test pressure was induced [22].

Creativity was assessed in three ways: DT (by using two different association tasks), CT and CSE. For DT, first, we asked participants to draw a picture of their favorite place (IMT). This was realized using a paint plug-in. Two raters experienced in AUT ratings did this evaluation following the Consensual Assessment Technique [23]. Each originality score is the mean of the individual idea scores. For originality, two scores from both raters, from 1 = low to 5 = very high, were assigned to the participants. Elaboration, as the extent of details shown in the picture, was assigned on a scale from 1 = low, to 5 = very high. Inter-rater reliability was high, with $ICC_{Originality}$=.84, with a 95% CI from .78 to .88 ($F(137)=11.59$, $p \leq .001$) and $ICC_{Elaboration}$=.80, with a 95% CI from .73 to .86 ($F(137)=9.17$, $p \leq .001$). The Alternate Uses Task (AUT) [24] was applied for the second creative association task, asking participants what they could think of to do with a brick. Responses were collected, and two different scores were derived: fluency, as the number of answers, and originality again, from 1 = low, to 5 = very high. All answers from all participants were evaluated separately to exclude influences of the idea frequency. Again, the same two raters assessed the originality scores. Their agreement was high with an intra-class correlation (ICC) of .87 with a 95% confidence interval (CI) from .81 to .91 ($F(99)=14.10$, $p \leq .001$).

CT was assessed using the RAT to measure individual ability to find the correct association for a closed problem space [25]. For each RAT item, three unrelated terms were presented, for which a matching term needs to be found. Only one word fits correctly. For example, *soda* would be the matching term to the triad *fountain / baking / pop.* The sum of correct responses out of 20 total items represented the overall RAT score. The items were selected to follow a normal distribution of difficulty based on the evaluation from https://www.remote-associates-test.com. The items were presented for 20 seconds, respectively, with automated page redirects to the next one. Based on two validation studies, the Spearman-Brown reliability revealed high measuring precision, with r = .92 and r = .91 [26]. Creative self-efficacy (CSE) was assessed using the three items of

Tierney and Farmer [16] with a Likert scale. A sample item would be "I have confidence in my ability to solve problems creatively". In prior studies, the scale reached a medium to high Cronbach's $\alpha \in [.74, .81]$ [27], which is in line with the reliability reached in this sample, $\alpha = .76$.

The online course was administered on the online participant recruitment platform *Prolific*. Participants were selected with age above 18, a minimum approval rate on Prolific of 95%, and fluent German language skills, as our online course was in German. Participants gained approximately 10$/h to complete the course. On average, the course took 61 minutes to master (SD = 20.3). Participants were primarily from Poland (43%) and Germany (15.5%). The mean age of the participants is 23.18 years (SD = 5.0, ranging from 18 to 50). 93.5% of the participants were students. Overall, 200 participants took part in the online course, of which 53% were male.

For deriving features, we used two tables of the Moodle database, "logstore_standard_log" for general trace data (T) and "hvp_xapi_results" for performance data (P) of interactive tasks. Finally, we had 59 features, consisting of 48 trace features and 11 performance-based features. Trace features were derived by 24 pages that contained the learning content, and we measured the time that learners spent on that page plus whether pages were accessed multiple times. In sum, we had 48 trace features. The performance features consisted of different interactive H5P tasks, where the learners had to find the correct solutions.

Next, we normalized the data, examined a Linear Discriminant Analysis (LDA) for feature dimension reduction, and trained this classifier with the collected data. To understand how well this model performs, we used cross-validation with 5 folds (5f-CV). Thus, we trained the model with four splits and tested it with the remaining one. Finally, we got a prediction on data that has not been used for training, and we could see how well the model performs with unseen data. The k-nearest-neighbor (KNN) was used as a non-parametric approach to predict all creativity metrics based on the surrounding classes in the feature space. The number of neighbors and balance of weights were optimized to get an optimal result. Next, a Random Forest Tree (RF) was trained to make predictions for the creativity metrics. First, it was trained with the original data as normalization does not affect RFs, and the resulting accuracy cannot be optimized by that in general. Then, we applied the LDA for feature reduction (not for the classification) and used the reduced features as inputs to train the RF. KNN and RF with reduced features were trained separately with trace data "T" (containing all click data without interactive tasks) and the performance data "P" (including the interactive parts only). The comparison of both accuracies in 5f-CV shows the differences based on the selected data.

IV. Results

In our study, some participants skipped some of the creativity tests. Thus, the remaining dataset differs in the number of participants: IMT: 138, AUT: 156, CSE: 200, RAT: 200 (125 found at least one correct answer).

All trace and performance data points were transformed using the Quantile Transformation (QT) that transfers all features to the standard probability distribution [28], except for the labels that have to be predicted, which were not transformed to remain the original values. The transformation was necessary as derived features were not always normally distributed and this procedure transforms the data for optimal usage. QT is used for non-linear transformations. As many features are skewed, this method centers the mean value of the data to 0 and results in a standard deviation of 1. The LDA assumes a Gaussian distribution of the input variables, which can be fulfilled with the transformation. Alternatively, a Box-Cox could be used. In the case of IMT and AUT, all prediction models were separately trained for the two raters to avoid working with means where classes could sometimes not be clearly determined as their mean value could be between two classes, which results in new classes that are under-represented in the overall dataset. The results show that the models perform similarly, with the same order of magnitude for all results (Table 1). The LDA reduced 59 features to 4. We can see that the LDA performs badly as a single predictor in cross-validation, but the results are better than guessing.

For the KNN, no assumptions were made for the decision boundary, and it dominated the LDA when the decision boundary was highly non-linear, which was the case in our experiments. Only in the IMT elaboration metric, the LDA performed better. However, this can be a coincidence. Compared to all metrics, the KNN performed better than the LDA. The RF without optimized input performed worst. Sometimes, the cross-validation accuracy was the worst for all experiments. In contrast, using QT for creating normally distributed data, making a feature reduction with the LDA first, and using the transformed input features to train the RF worked best for the IMT and AUT, and it outperformed the other results. For CSE and RAT, the LDA+RF combination performed equal to the KNN.

To understand the accuracy and the practicableness of the results, a closer investigation to the number of classes for each creativity test was necessary. The originality, elaboration, and CSE [1-3] metrics were based on a Likert scale (5 points). Thus, we had 5 classes, which were predicted. Predicting just a class by choice, we got an accuracy of .2. All predictions reached ~.4 without optimizations, sometimes even more than .5, and the optimal combination reached ~.7, which is comparable with prediction tasks in educational data mining [4]. This result shows that the trace and performance data can be used to predict creativity metrics. To summarize, originality and elaboration could both be predicted. In our experiments, the accuracy of predicting IMT scores was slightly better than for AUT. The fluency metric of the AUT could not be predicted in our experiments. The result was close to coincidence. The RAT's performance was also bad due to many data points where no learner was aware to find at least one correct solution. We transformed the RAT results into three classes to have a uniform distribution. The resulting accuracy was close to .33, which is close to guessing. Thus, we can conclude that the RAT for convergent thinking and fluency in AUT cannot be predicted based on our online course's trace and performance data. CSE items could be predicted with an accuracy of ~.45 in cross-

TABLE I. RESULTS FOR IMT, AUT, CSE (THREE ITEMS SEPARATELY), AND RAT IN 5F-CV USING LDA, KNN (IMT: NEIGHBORS = 9, WEIGHTS = DISTANCE-1, AUT/CSE/RAT: NEIGHBORS = 29, UNIFORM WEIGHTS), AND RF

	DT							**CSE**			**CT**
	IMT				***AUT***						
	Orig. (1)	**Orig. (2)**	**Elab. (1)**	**Elab. (2)**	**Orig. (1)**	**Orig. (2)**	**Fluency**	**Item 1**	**Item 2**	**Item 3**	**RAT**
LDA (T + P)	.40	.35	.47	.46	.45	.36	.19	.40	.41	.42	.25
KNN (T + P)	.51	.46	.42	.47	.46	.37	.24	.44	**.46**	.44	**.36**
KNN (T)	.42	.42	.44	.47	.49	.40	.21	.44	.45	.45	.34
KNN (P)	.45	.46	.38	.42	.42	.36	.23	.44	.44	.45	.31
RF (T + P)	.44	.35	.35	.39	.43	.31	.13	.28	.27	.29	.13
LDA → RF (T + P)	**.70**	**.69**	**.74**	**.72**	**.66**	**.64**	**.50**	**.45**	.44	**.48**	.34
LDA → RF (T)	.56	.58	.62	.63	.50	.46	.35	.41	.40	.46	.32
LDA → RF (P)	.50	.47	.42	.43	.41	.45	.22	.29	.31	.29	.25

Notes: DT = divergent thinking, CSE = Creative self-efficacy, CT = convergent thinking, IMT = image test, AUT = alternate uses test, RAT = remote association test, Orig = originality, Elab = elaboration,

validation. Compared to the other metrics, this is a low value, and originality and elaboration perform much better.

We also investigated in training the KNN and the LDA+RF with trace data (without performance information, T) and with performance data (P) separately to understand whether this subset of features could be used to address data economy. For originality in IMT, there was no clear result on which of both variants performed best as in KNN, P was slightly better, and for LDA+RF, T resulted in the optimal accuracy. For Elaboration, T performed best in both approaches. For originality in AUT, T also performed best in both approaches. All in all, using the overall feature set performed much better than using a subset, and the accuracy in cross-validation increased at least by 10%.

V. DISCUSSION

Based on logfile data of an online course, we aimed to predict several creativity scores, from which originality and elaboration of DT could be successfully predicted. It is important to note that being able to predict individual creativity scores does not mean that we found causal inferences [29]. Thus, an explanation cannot be based on features with high importance in prediction. In contrast, focusing on trace or performance data only, the trace data performs slightly better. Comparing the results for all scores using T or P only (Table 1), using the subset performs worse in relation to using all features. None of the separate feature sets is in the range of the best accuracy where we use all features. T is recommended if a lower accuracy is acceptable as it mainly performs better than P or is quite similar. One reason for the lower performance of P could be that data often consists of binary information (e.g. whether the learner solved the task correctly) and T consists of more enriched data (e.g. the time participants spent with the learning material). From the data economy perspective, this is an important finding. We principally could exclude performance data with just a minor loss of accuracy. However, both feature sets contain information that must be used for the best achievable accuracy in prediction.

Using normalized data does not affect RFs in general. However, within the experiments, we could see that a feature reduction is highly interesting as it influences the accuracy. For a real-world scenario, it is essential that the data transformation also has to be done with data of new participants before a prediction is possible. This is an additional effort as the data has to be transformed as it has been done before training. It can be discussed whether a prediction accuracy of ~.7 is sufficient. The result is a common finding in online learning, the accuracy is often not better [7]. For personalizing online courses, a rough range is acceptable. We used 5 classes, but in a real-world scenario, creating two task versions seems realistic in terms of the effort required in a real-world scenario. Thus, having 2 or 3 classes (low/high or low/medium/high) is sufficient. Having a value of ~.7 for five classes, 2 or 3 classes will increase the accuracy as we merge the data. Besides, as the human evaluations' agreement was also in the range of ~.8, we conclude that ~.7 is an acceptable result. This is a legitimate comparison [30].

As the Componential Theory from Amabile [23] shows, creative performance is hugely influenced by the individual's motivation to engage with creative problems and tasks. This motivation is highly influenced by situational conditions and the individual perceived capability of creative abilities. Such self-evaluations and the actual performance can be best improved through individual, adequate, and in tendency, positive feedback. Online courses integrating the assessment of such creativity-associated self-conceptions and performances could present input based on such individual needs. Further, learners' motivation might be improved to deeply engage in problem-solving tasks by providing more open or closed creative tasks, primarily based on DT, as this can be best predicted. This would enhance the scope of course adaptations beyond the topics typically associated with creativity, as the adaption based on DT is relevant for all courses that contain problem-solving tasks. However, theoretically, it could also be argued that presenting studies outside of a person's comfort zone would push the individual to enhance creative efforts [31]. Another future scenario could be the individualization to train creative abilities directly. Those facets of creative performance with the highest potential for improvement can be explicitly taught and supported by specific creativity methods (e.g. [32]).

Overall, this study can serve as a preparation for further individualization approaches to online learning courses. Knowing the learner's creative performance level could be used to find the right fit between the type of task, task difficulty, and the subjective feeling of potential mastery. Such a fit leads to the greatest engagement with the learning content, leading the learner to become wholly immersed and enjoy the task [33]. Such a state, especially as positive emotions are in place, is ideal for learning and the feeling of competence and control. At least improving learning conditions should be the goal of current online courses to exploit their full potential.

VI. Conclusion

In this paper, we examined the prediction of creativity scores based on logfile data of an online course. Combining an LDA with a RF outperforms other models and works best in our experiments. Divergent thinking, represented by originality and elaboration, can be predicted with an accuracy of ~.7. The prediction accuracy agrees with human evaluations' (ICC: ~.8). Thus, we can conclude that the prediction performs almost similarly to human judgment. Fluency cannot be predicted well. Convergent thinking, measured by RAT, cannot be predicted appropriately. The perceived creative self-efficacy achieves an accuracy of ~.45, and the combination of the LDA with the RF does not influence the result remarkably. All in all, we got promising results for two divergent thinking scores that should be addressed in further experiments.

Acknowledgment

This work was supported by the German Federal Ministry of Education and Research (BMBF), grant number 16DII127 (Weizenbaum-Institute). The responsibility for the content of this publication remains with the authors.

References

[1] D. K. Srivastava and H. Shah, "Do learning styles of undergraduate and postgraduate students in B-schools differ? Insights and implications," in *Journal of Education for Business*, 2021, pp. 1-8.

[2] J. Rafner, M. Biskjær, B. Zana, S. Langsford, C. Bergenholtz, S. Rahimi, A. Carugati, L. Noy and J. Sherson, "Digital Games for Creativity Assessment: Strengths, Weaknesses and Opportunities," in *Creativity Research Journal*, 2021, pp. 1–27.

[3] B. Dietz-Uhler and J. E. Hurn, "Using learning analytics to predict (and improve) student success: A faculty perspective," in *Journal of interactive online learning 12(1)*, 2013, pp. 17-26.

[4] 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.

[5] B. Naik and S. Ragothaman, "Using neural networks to predict MBA student success," in *143-150*, Project Innovation (Alabama), 2004, pp. 143-150.

[6] M. Kosinski, D. Stillwell and T. Graepel, "Private traits and attributes are predictable from digital records of human behavior," Free School Lane, The Psychometrics Centre, University of Cambridge, Cambridge CB2 3RQ United Kingdom; and Microsoft Research, Cambridge CB1 2FB, United Kingdom, Proc Natl Acad Sci USA, 2013, pp. 5802-5805.

[7] B. Albreiki, N. Zaki and H. Alashwal, "A Systematic Literature Review of Student'Performance Prediction Using Machine Learning Techniques," in *Education Sciences, 11(9), 552*, 2021.

[8] P. Kerr, "Adaptive learning," in *ELT Journal 70(1)*, 2015, pp. 88–93.

[9] V. P. Glaveanu and e. al., "Advancing Creativity Theory and Research: A Socio-cultural Manifesto," in *The Journal of Creative Behavior*, 2019.

[10] M. A. Runco and S. Acar, "Divergent Thinking as an Indicator of Creative Potential," in *Creat. Res. J., (24), no. 1*, 2012, pp. 66–75.

[11] R. J. Sternberg, "Thinking Styles," Cambridge University Press, 1999.

[12] G. Scott, L. E. Leritz and M. D. Mumford, "The effectiveness of creativity training: A quantitative review," in *Creat. Res. J. 16(4)*, 2004, pp. 361–388.

[13] L.-F. Zhang, "Thinking Styles and Modes of Thinking: Implications for Education and Research," in *The Journal of Psychology 136(3)*, 2002, pp. 245–261.

[14] L.-F. Zhang, "Revisiting the Predictive Power of Thinking Styles for Academic Performance," in *The Journal of Psychology 138(4)*, 2004, pp. 351–370.

[15] M. Yagci, "Blended Learning Experience in a Programming Language Course and the Effect of the Thinking Styles of the Students on Success and Motivation," in *Turkish Online Journal of Educational Technology - TOJET 15(4)*, 2016, pp. 32–45.

[16] P. Tierney and S. M. Farmer, "Creative self-efficacy: Its potential antecedents and relationship to creative performance," in *Academy of Management Journal 45, no. 6*, 2002, pp. 1137–1148.

[17] C. Byrge and C. Tang, "Embodied creativity training: Effects on creative self-efficacy and creative production," in *Thinking Skills and Creativity 16*, 2015, pp. 51–61.

[18] X. Wang, K. Lu, M. A. Runco and N. Hao, "Break the 'wall' and become creative: Enacting embodied metaphors in virtual reality," in *Consciousness and cognition 62*, 2018, pp. 102–109.

[19] D. Ismunandar, F. Gunadi, M. Taufan and D. Mulyana, "Creative thinking skill of students through realistic mathematics education approach," in *Journal of Physics: Conference Series 1657*, 2020, pp. 012054.

[20] M. Hartnett, "The importance of motivation in online learning," in *Motivation in online education*, Singapore, Springer, 2016, pp. 5-32.

[21] K. Sodhi, M. Luthra and D. Mehta, "Yerkes-dodson law for flow: A study on the role of competition and difficulty in the achievement of flow," in *International Journal of Education and Management Studies 6(1)*, 2016, pp. 95.

[22] T. L. Belcher, "Effect of Different Test Situations on Creativity Scores," in *Psychological Reports 36(2)*, 1975, pp. 511–514.

[23] T. M. Amabile, "Componential theory of creativity," in *Harvard Business School 12(96)*, 2012, pp. 10.

[24] P. Christensen, J. Guilford, R. Merrifield and R. Wilson, "Alternate uses test," Beverly Hills, CA, Sheridan Psychological Service, 1960.

[25] E. M. Bowden and M. Jung-Beeman, "Normative data for 144 compound remote associate problems," in *Behavior Research Methods, Instruments, & Computers 35(4)*, 2003, pp. 634–639.

[26] S. Mednick, "The associative basis of the creative process," in *Psychological review 69(3)*, 1962, pp. 220.

[27] P. Tierney and S. M. Farmer, "Creative self-efficacy development and creative performance over time," in *Journal of Applied Psychology 96, no. 2*, 2011, pp. 277.

[28] P. McCullagh and M. F. Tresoldi, "A likelihood analysis of quantile-matching transformations," 2021, pp. 247–251.

[29] P. Spirtes, C. Glymour and R. Scheines, "Causation, Prediction, and Search," in *Adaptive Computation and Machine Learning*, MIT Press, 1993.

[30] S. Rüdian and N. Pinkwart, "Using Data Quality to compare the Prediction Accuracy based on diverse annotated Tutor Scorings," in *International Conference on Educational Data Mining*, 2021.

[31] R. Sunley, L. Harding and J. Jones, "Realising creativity in management education: Putting student energy into action," in *The International Journal of Management Edu. 17(2)*, 2019, pp. 172–181.

[32] M. Michalko, "Thinkertoys: A handbook of creative-thinking techniques," Ten Speed Press, 2010.

[33] M. Csikszentmihalyi, S. Abuhamdeh and J. Nakamura, "Flow," New York, Harper & Row, 1990.