Education and Science

Vol 47 (2022) No 210 121-137

Original Article

Predicting Geospatial Thinking Ability for Secondary School Students Based on the Decision Tree Algorithm in Mainland China

Shumin Xie [1], Siying Zeng [2], Lu Liu [3], Huimin Wei [4], Yanhua Xu* [5], Xiaoxu Lu* [6]

Abstract

Predicting secondary school students' geospatial thinking ability can provide targeted guidance for teachers. To date, few scholars have focused on predicting students' geospatial thinking ability. In this paper, we address this gap by constructing a prediction model based on the decision tree algorithm, to predict the geospatial thinking ability of secondary school students. A total of 1029 secondary school students were surveyed using the Spatial Thinking Ability Test, the Students' Geography Learning Status Questionnaire, and the Middle Students Motivation Test. Our model indicates that geospatial thinking ability can be predicted by nine factors, in order of importance: academic achievement in geography, geography learning strategy, geography classroom environment, gender, learning initiative, learning goals, extra-curricular time spent learning geography, ego-enhancement drive, and interest in learning geography. The model accuracy is 81.25%. Specifically, our study is the first to predict geospatial thinking ability. It provides a tool for teachers that can help them identify and predict students' geospatial thinking ability, which is conducive to designing better teaching plans and making adjustments to the curriculum.

Keywords

Geospatial thinking ability
Decision tree
Prediction
Model
Teaching

Article Info

Received: 12.07.2020
Accepted: 03.30.2022
Online Published: 04.29.2022

DOI: 10.15390/EB.2022.10367

[1] East China Normal University, Faculty of Education, College of Teacher Education, China, shuminxieedu@foxmail.com
[2] East China Normal University, Faculty of Education, College of Teacher Education, China, 51204800016@stu.ecnu.edu.cn
[3] East China Normal University, Faculty of Education, College of Teacher Education, China, 51204800156@stu.ecnu.edu.cn
[4] East China Normal University, Faculty of Education, College of Teacher Education, China, 51204800167@stu.ecnu.edu.cn
[5] * **Corresponding author:** Capital Normal University, College of Resource Environment and Tourism, China, yanhuaxuedu@foxmail.com
[6] * **Corresponding author:** East China Normal University, Faculty of Education, College of Teacher Education, China, xxlu@geo.ecnu.edu.cn

Introduction

Since the National Research Council published a report entitled *Learning to think spatially*, spatial thinking has become a popular topic of inquiry in the field of education. The report defined spatial thinking as comprising the concepts of space, tools of representation, and processes of reasoning (National Research Council, 2006). It argued that spatial thinking is important in daily life, in the workplace, and for academic achievement, and that it should therefore be taught in schools. In response, an increasing number of scholars and educators began researching the merits of spatial thinking. For instance, Garcia De la Vega (2019) listed situations in which people used spatial thinking in their daily activities, such as observing weather conditions, determining navigation directions, going to work, traveling, and moving. Montello, Grossner, and Janelle (2014) emphasized the importance of spatial thinking in many professional fields, such as geology, architecture, and surgery. In addition, scholars have conducted empirical studies analyzing the influence of spatial thinking on academic achievements in the fields of science, technology, engineering, and mathematics (Atit, Uttal, & Stieff, 2020; Stieff et al., 2018; Uttal, Miller, & Newcombe, 2013; Wai, Lubinski, & Benbow, 2009).

In scholastic education, the discipline of geography explicitly deals with spatial thinking. Recognized as a space science (Hartshorne, 1958), geography is conducive to the cultivation of spatial thinking. Geographic thinking includes the understanding of spatial concepts and their relationships, explaining high-level correlations derived from geographic concepts, and theorizing the correlations (Metoyer & Bednarz, 2017). Additionally, geographic knowledge is considered to be the product of geographic thinking and reasoning (Golledge, 2002), which is developed around spatial location, spatial distribution, and spatial relationships (Catling, 1978). Hence, scholars have argued that the discipline of geography could and should teach students spatial thinking skills (Anthamatten, 2010; Hilman & Mainaki, 2020; Kim & Bednarz, 2013; Yani, Mulyadi, & Ruhimat, 2018).

The type of spatial thinking taught within the geography discipline is called geospatial thinking (Verma, 2015). Geospatial thinking involves the use of cognitive skills to combine and transform various forms of knowledge (Painho, Santos, & Pundt, 2010). Moreover, geospatial thinking is defined as the ability to use spatial concepts and representation tools to carry out reasoning for specific geographic problems, and coming up with methods for solving the problems (Lobben & Lawrence, 2015; National Research Council, 2006; Verma, 2015).

In the literature on geography education, scholars have studied and developed various measurement and assessment tools for geospatial thinking ability. The measurement of spatial thinking ability in the context of geography or earth science can be traced back to the Geologic Spatial Ability Test (GeoSAT). This test requires students to imagine geologic maps and cross sections, and to draw them (Kali, Orion, & Mazor, 1997). Several years after the development of the GeoSAT, spatial thinking tests for geography education appeared. However, their quality was not verified until the emergence of the Spatial Skill Test (SST) developed by Lee and Bednarz (2009). In order to measure spatial thinking ability in the context of geography and earth science more accurately, Lee and Bednarz (2012) developed the Spatial Thinking Ability Test (STAT) based on the SST, which was later used in many academic studies (e.g., Collins, 2018; Flynn, 2018; Tomaszewski, Vodacek, Parody, & Holt, 2015). In addition, scholars have developed some tools for assessing geospatial thinking ability. For instance, Huynh and Sharpe (2013) developed a tool for assessing geospatial thinking, which mainly focuses on students' understanding of spatial relations within a geographic context.

However, fewer studies have focused on predicting people's geospatial thinking ability. Specifically, predicting students' level of geospatial thinking can provide effective guidance for geography teachers on how to help students develop these skills. Abu (2016) attached great importance to the prediction of academic success, claiming that educational institutions could make additional efforts in helping students in their studies and academic success. Geospatial thinking ability, as one of the components of geography learning, is no exception. Predicting and identifying which students have lower geospatial thinking ability can enable teachers to give them more targeted guidance, which can help students improve their geospatial thinking ability.

It is necessary to fill the gap between the assessment and prediction of geospatial thinking ability. Thus, we pose the following research question: How do we predict students' geospatial thinking ability?

We need to search for predictors of geospatial thinking ability. First, scholars found that the geospatial thinking ability of male and female students was significantly different (Shin, Milson, & Smith, 2016; Tomaszewski et al., 2015). Thus, in our study, we hypothesized that gender could predict geospatial thinking ability (H1). Second, geography learning interest, defined as one’s interest to explore geographical knowledge, was also found to be important in influencing geospatial thinking ability (Wakabayashi, 2015; Wan et al., 2017). We hypothesized that geography learning interest could predict geospatial thinking ability (H2). Third, academic achievement in geography, defined as the grades obtained by students in geography class, is also closely related to geospatial thinking ability (Wan et al., 2017). We considered academic achievement in the field of geography as key predictor of geospatial thinking ability (H3). Forth, it is believed that spatial thinking ability can be activated by learning about geographic topics (Yani et al., 2018), while students’ learning results were influenced by learning time, classroom environment and learning strategy (Berberoğlu & Demircioğlu, 2000; Chan, Wong, & Lo, 2012; Ergene, 2011). We hypothesized that geospatial thinking ability could be predicted by extra-curricular time spent learning geography (H4), the geography classroom environment (H5), and the geography learning strategy (H6). Specifically, extra-curricular time spent learning geography is the time students spent on geography studies out of the school classroom. Geography classroom environment refers to the classroom environment created by the interaction between teachers and students, including students' engagement in the classroom, teachers' support, and classroom discipline. Geography learning strategy is a processing strategy adopted by students to improve the efficiency of geography learning. According to Entwistle, McCune, and Walker (2001), there are two main learning strategies. One is deep processing, where students try to understand the meaning of learning materials and explore the relationship between old and new knowledge. The other is surface processing, which means that students learn through automation, memorization, and repetition. Fifth, students' motivation and desire for academic learning is an intrinsic factor that directly helps them learn. According to Ausubel, Novak, and Hanesian (1978), high academic achievement can be attributed to three internal driving forces (Shi, 1994). The first factor is cognitive drive, which manifests itself as the interest in learning. People learn in order to satisfy their curiosity, seek truth, and satisfy their thirst for knowledge. The second factor is the ego-enhancement drive, where students seek to increase their status, honor, or self-esteem through learning. The third factor is the affiliated drive, which happens when students are praised and recognized by parents and teachers for their academic achievements. This kind of motivation usually occurs in early childhood. For secondary school students, cognitive drive and ego-enhancement drive are the main factors that motivate them to learn and are important in shaping learning outcomes (Shi, 1994). Therefore, we assumed that cognitive drive (H7) and ego-enhancement drive (H8) could predict geospatial thinking ability. Sixth, although students may want to learn and perform well academically, they may not be willing to take the initiative to learn. Learning initiative refers to the positive psychological state of students when they are engaged in learning activities, which is manifested as learning in an active and willing manner (Lin, Yang, & Huang, 2004). Learning goals, defined as students' learning requirements for themselves, also influence student academic achievement (Lozano, Uzquiano, Riobo, Malmierca, & Blanco, 2011). Hence, we hypothesized that learning initiative (H9) and learning goals (H10) can help predict geospatial thinking ability. The conceptual framework of the study is shown in Figure 1.

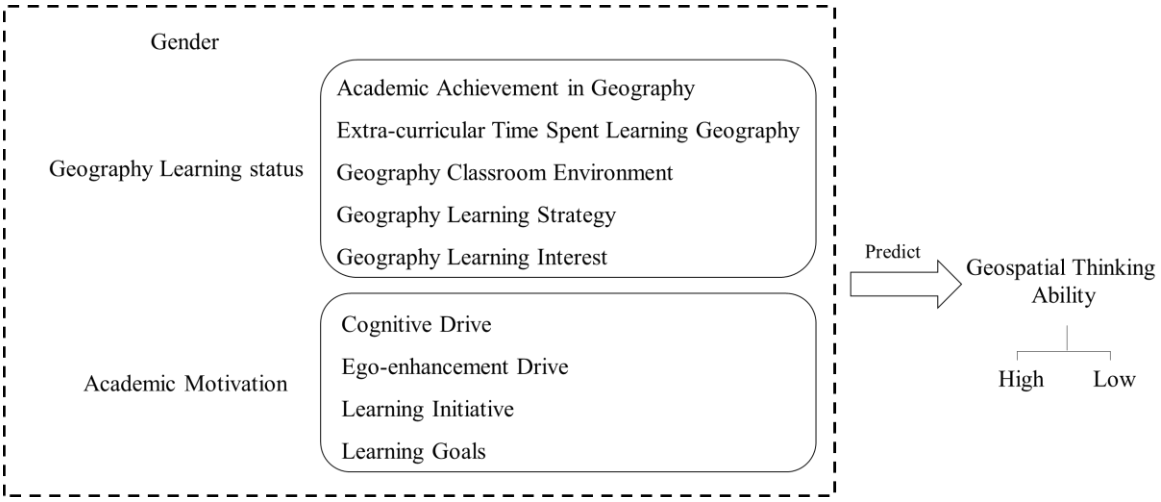

Figure 1. Hypothetical predictors of geospatial thinking ability

Method

Participants

Our study was conducted in 27 classes across nine secondary schools in the eastern, middle, and western regions of China. Using a stratified sampling method, we selected participants through random sampling in the eastern, middle, and western regions of China, according to regional differences. Specifically, we selected the provinces of Guangdong, Fujian, and Shanghai in the eastern region; Inner Mongolia, Anhui, and Jiangxi in the middle region; and Tibet, Yunnan, and Gansu in the western region. In each province, we selected one local school, based on the principle of intergroup homogeneity. In each school, we selected three geography classrooms in Senior Two: one with high academic performance, one with medium performance, and one with poor performance. In China, the students in the Senior Two geography class have acquired the geography knowledge required at the secondary level. Finally, a total of 1199 students agreed to participate in the study and filled in questionnaire. Most of the participants were 16 years old, and a few were 15 or 17 years old. We received responses from all 1199 students. After excluding 166 respondents who failed to pass the lie detection procedure and 4 respondents whose answers were incomplete, the final sample consisted of 1029 students with an efficiency rate of 85.82%. Of the 1029 respondents, 362 were from the eastern region, 388 were from the middle region, and 279 were from the western region; 331 were male and 698 were female. Students entering the geography class tend to choose liberal arts subjects for the college entrance examination. In China, liberal arts students are mainly females. Therefore, the number of females in the sample is greater than males.

Materials

The Spatial Thinking Ability Test

We used the Spatial Thinking Ability Test (STAT) developed by Lee and Bednarz (2012) to measure students' geospatial thinking ability. The test has 16 multiple-choice items with questions on geographical situations, which are based on eight components of spatial thinking. There is only one correct answer for each item. One point is given for a correct answer, and the maximum score is 16. The STAT is widely used and it is considered to be one of the best tools for measuring spatial thinking ability within a geographical context (Bednarz & Lee, 2019). In China, Wan et al. (2017) translated the STAT into Chinese and tested it in a middle school. After translating this Chinese version of the STAT back to English, we found that the questions were similar to the original version, indicating that the translation did not distort the original meaning. In our study, the Cronbach's alpha of the STAT was 0.717.

The Students' Geography Learning Status Questionnaire

The Students' Geography Learning Status Questionnaire was adapted from the Student Common Part Questionnaire by the Program for International Student Assessment (Organisation for Economic Co-operation and Development, 2015) that asked students to report information about themselves, their families, and their attitudes towards learning. The adaptation covered five test topics, namely academic achievement, extra-curricular learning time, classroom environment, learning strategy, and learning interest, all of which were asked in the context of the geography discipline. Academic achievement in geography was assessed using students' response to the question "Generally, what is the range of your geography test scores?" Extra-curricular time spent learning geography was measured by students' response to "In addition to geography classes at school, how much time do you spend on geography studies (including homework, extra-curricular tutoring or other personal arrangements) per week?" Geography classroom environment was measured based on 21 items (alpha=0.894), including "How often do these things happen in your geography classes?" and by responses to statements such as "Students don't listen to what the teacher says." The question "How often do these things happen to you?" aimed at assessing geography learning strategy (alpha=0.753), and contained 11 items, including the statement "I'll make sure I understand the logic of geography." Finally, there the question "Are the following items in line with your reality?" aimed at assessing interest in learning geography (alpha=0.756), and also contained 11 items, one of which being "I like to talk about geography with others."

Middle Students Motivation Test

The Middle Students Motivation Test (MSMT) was designed by Zheng (1994). It was tested on 560 students at the time it was developed, and the scale received a 95% correctness rate. We used 20 questions from the MSMT to assess middle school students' cognitive drive, ego-enhancement drive, learning initiative, and learning goals. The test requires students to assess whether the description of the items is consistent with their actual situation, and then to select "yes" or "no". An example of a test item assessing cognitive drive is "I often feel that there is nothing to learn from the basic knowledge of textbooks. Only profound theories and classic works are interesting." An example of the ego-enhancement drive assessment is "I am eager to improve my academic performance in a short period of time". An example of learning initiative assessment is "I feel tired and want to sleep when it comes to reading". This is also a reverse problem. An assessment example of learning goals is "I'm always struggling to achieve several learning goals at the same time." In our research, the Cronbach's alpha of the MSMT was 0.670.

Lie Detection Questions

Lie detection questions were used to detect whether the respondents answered the questionnaire carefully and whether they were inclined to lie when filling out the questionnaire. There were five lie detection questions in our questionnaire. One example is, "I went to Antarctica via Ushuaia in South America." The student who selected "yes" was considered to have failed the lie detection test.

Design

This is a quantitative study aimed at revealing the predictors of geospatial thinking ability. We used the questionnaire method to obtain students' performance in geospatial thinking ability and ten predictors. We submitted the questionnaire materials and explained our survey design to the Ethics Review Committee of the East China Normal University and to the directors of sample schools. They held meetings to discuss research ethics and survey requirements, respectively. After obtaining their consent, we conducted a survey questionnaire with the students.

We chose the data mining method to process the survey data, which is useful for uncovering patterns that can help predict students' academic performance (Francis & Babu, 2019). Among the numerous data mining methods, the Decision Tree, k-Nearest Neighbor, Neural Network, Naïve Bayes, and Support Vector Machine are commonly used and are considered suitable for predicting students' academic performance (Ramaswami, Susnjak, Mathrani, Lim, & Garcia, 2019; Xu, Wang, Peng, & Wu, 2019).

In this paper, we selected the decision tree to predict students' geospatial thinking ability for three reasons: (i) the decision tree model has been widely used to predict students' academic performance and learning behavior (Asif, Merceron, Ali, & Haider, 2017; Chen et al., 2019; Hamoud, Hashim, & Awadh, 2018; Suguna, Shyamala Devi, Bagate, & Joshi, 2019); (ii) the decision rules are easy to understand. To be specific, the decision tree analysis established classification rules through training samples, with which new samples can be classified (Han, Kamber, & Pei, 2012). The output of the analysis is a top-down diagram that is easy to understand and explain (Tan, Steinbach, & Kumar, 2016). In the diagram, a decision tree is composed of a root node, several internal nodes, and several leaf nodes. The root node and internal node represent the corresponding test conditions (or criteria of classification), while the leaf node represents the final output. We can infer rules according to the tree structure formed by each node (Mitchell, 1997); and (iii) the decision tree algorithm has a good tolerance for multicollinearity and can deal with the complex relationship between predictors. In addition, the classification decision tree is used when the predictive variable is categorical, and the regression decision tree is suitable for continuous predictive variables (Miguéis, Freitas, Garcia, & Silva, 2018). In this study, we seek to identify students' geospatial thinking ability, which we classified as high or low. As such, we used the classification decision tree algorithm to construct our prediction model and analyzed the importance of each factor in predicting geospatial thinking ability.

Data Analysis

We used SPSS 22.0 to conduct descriptive statistical analysis and Modeler 18.0 for the decision tree. First, the descriptive statistical analysis was mainly implemented to analyze the frequency statistics and concentration trends regarding students' level of geospatial thinking ability and the predictors. Second, the decision tree analysis was conducted with the C5.0 algorithm to construct the prediction model for geospatial thinking ability. We chose the C5.0 algorithm because it was an extension of the ID3 algorithm and of the C4.5 algorithm proposed by Quinlan (1986, 1992) and by Witten, Frank, and Hall (2011), which is not only suitable for big data but also has faster running speed and better predicting ability (Xiong, 2011).

Procedure

Data Coding

We divided the samples into two groups, high geospatial thinking ability and low geospatial thinking ability, with 60% as the cutoff point. As for the predictors, we transformed the nominal or continuous variables for students' geography learning status, learning motivation, and gender into binary variables (see Table 1) according to certain criteria.

Table 1. Variable coding and their descriptive statistics

Variable	Coding	Number	Proportion
Geospatial thinking ability	0=low	200	19.44%
	1=high	829	80.56%
Gender	0=female	698	67.83%
	1=male	331	32.17%
Academic achievement in geography	0=low	335	32.56%
	1=high	694	67.44%
Extra-curricular time spent learning	0=short	463	45.00%
geography	1=long	566	55.00%
Geography classroom environment	0=bad	117	11.37%
	1=good	912	88.63%
Geography learning strategy	0=surface	374	36.35%
	1=deep	655	63.65%
Geography learning interest	0=low	524	50.92%
	1=high	505	49.08%
Cognitive drive	0=good	991	96.31%
	1=bad	38	3.69%
Ego-enhancement drive	0=good	609	59.18%
	1=bad	420	40.82%
Learning initiative	0=good	908	88.24%
	1=bad	121	11.76%
Learning goals	0=low	635	61.71%
	1=high	394	38.29%

The Construction of the Decision Tree

When constructing the decision tree, the descending speed of information entropy is used to determine the best branch variable and segmentation threshold. Information entropy represents the degree of impurity of a data set and is defined based on Mitchell (1997) as:

$$Entropy(D) = -\sum_{k=1}^{m} P_k log_2 P_k \quad (1)$$

D is a training data set with sample size m and P_k is the probability of each class of samples. The information gain ratio is used to measure the information entropy difference of data sets under different classification methods. If we select variable C to divide the data set D into n subsets, then the information gain ratio is defined based on Quinlan (1996) as:

$$Gain\ ratio\ (D,C) = \frac{Entropy(D)-Entropy(D|C)}{Entropy(C)} \quad (2)$$

The C5.0 algorithm selects the attribute with the maximum information gain ratio as the splitting point, establishes several branches according to the value of this attribute, and obtains some subsets. This selection process is repeated until the final subset contains only the data of the same category, to perform the inductive classification for the data (Che, Liu, Rasheed, & Tao, 2011).

Pruning of the Decision Tree

The C5.0 algorithm uses the method of post-pruning to prune the leaves, layer by layer, from the leaf nodes. After the decision tree was constructed, the data set was recursive to each leaf node of the tree, according to the trained decision tree model. The mean square error of the data set with and without leaves was calculated. If the mean square error decreased after pruning, the node was cut off, otherwise it was retained (Quinlan, 2019).

Evaluation of the Decision Tree

We took 80% of the sample data (n = 821) as the training data and the remaining 20% (n = 208) as the test data. Whether the model constructed by the training data was suitable for new data was reflected by the test data. Model quality was evaluated based on accuracy, precision, and recall (Han et al., 2019). Accuracy refers to the proportion of correctly classified cases in relation to the total sample size. Precision refers to the prediction results, which indicates how many samples with positive prediction are real positive samples. Recall applies to the actual sample, showing how many positive examples in the sample are predicted correctly.

We evaluated the classification based on two additional indicators, the true positive rate (TPR) and the false positive rate (FPR) (Xiong, 2011). The TPR is the proportion of positive cases that were correctly predicted compared to the total number of positive cases, whose mathematical expression is similar to the recall procedure. The FPR is the proportion of negative cases that are wrongly predicted as positive cases in relation to the total number of negative cases. Taking the TPR as the Y-axis and the FPR as the X-axis, the receiver operating characteristic (ROC) curve was obtained. A larger area under the ROC curve (AUC) corresponds to a more accurate classification (Fawcett, 2006).

Results

Descriptive statistics

The descriptive statistics are summarized in Table 2. The forecast target, students' geospatial thinking ability, shows a good status. The mean value of geospatial thinking ability was 11.67 (with a standard deviation of 2.914), which is higher than 60% of the full score. This means that most students were in a state of high geospatial thinking ability. We then encoded each variable by assigning a value of 1 to cases with scores above 60% of the full score and a value of 0 to all other cases.

Table 2. Descriptive statistics

Variable	Full score	Mean value	Standard deviation	60% of the full score
Geospatial thinking ability	16	11.67	2.914	9.6
Academic achievement in geography	6	3.04	1.026	*
Extra-curricular time spent learning geography	5	2.78	1.006	3
Geography classroom environment	105	75.85	11.340	63
Geography learning strategy	55	34.75	5.959	33
Geography learning interest	55	32.60	5.467	33
Cognitive drive	5	0.82	0.845	3
Ego-enhancement drive	5	2.22	1.153	3
Learning initiative	5	0.83	1.236	3
Learning goals	5	2.01	1.458	3

*In the examination of students' geography academic achievement, as long as the students reported that their usual geography examination scores were above 60/100, they were considered to have high academic achievement in geography, and were assigned a value of 1.

Model Predicting Geospatial Thinking Ability

The decision tree model predicting geospatial thinking ability is shown in Figure 2. The prediction rules included in the model are as follows.

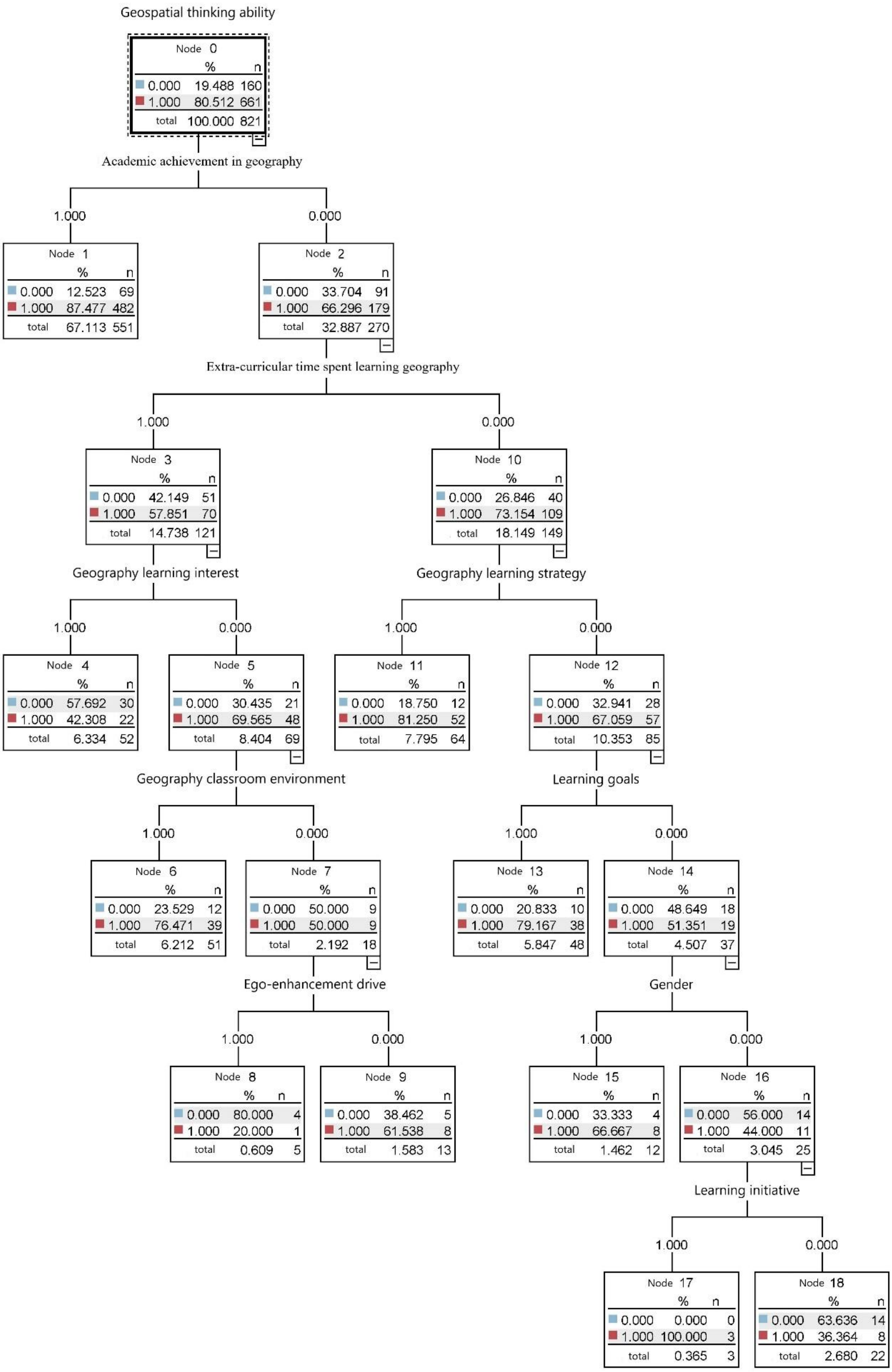

Figure 2. Prediction model for geospatial thinking ability

Academic achievement in geography is the first variable predicting geospatial thinking ability. Students with high achievement in geography were evaluated as having high geospatial thinking ability with an accuracy rate of 87.47%. Students with low geography academic achievement were predicted according to two branches under the extra-curricular learning time in geography.

Extra-curricular learning time in geography is the second variable that predicts geospatial thinking ability. For students who spent a lot of extra-curricular time learning geography, three variables helped predict their geospatial thinking ability. First, students with high interest in learning geography are evaluated as having low geospatial thinking ability (57.69%), while those with low geography learning interest were further evaluated according to the geography classroom environment. Second, if the classroom environment was good, students were classified as having high geospatial thinking ability (76.47%). Otherwise, they needed to behave with good ego-enhancement drive, so as to be identified as individuals with high geospatial thinking ability (61.54%). For students who spent little extra-curricular time learning geography, four variables were adopted to predict the geospatial thinking ability. First, students that had a deep learning strategy were classified as having high geospatial thinking ability (81.25%), while students using surface learning strategies need further evaluation. Second, if students set ambitious or more challenging learning goals, they were classified into the group with high geospatial thinking ability (79.17%). If students' learning goals were relatively at a low level, males were classified as having high geospatial thinking ability (66.67%) while females with low learning initiative were considered to have high geospatial thinking ability (100%).

The importance of the variables in the model, reflecting the contribution to the prediction, is shown in Figure 3. Among all the predictors, geography academic achievement was the most important. Geography learning strategy ranked second in order of importance, followed by the geography classroom environment. Gender, learning initiative, learning goals, extra-curricular time spent learning geography, ego-enhancement drive, and geography learning interest were found to be less important. It is worth mentioning that these variables were more accurate when predicting geospatial thinking ability than cognitive drive; however, this does not appear in the model.

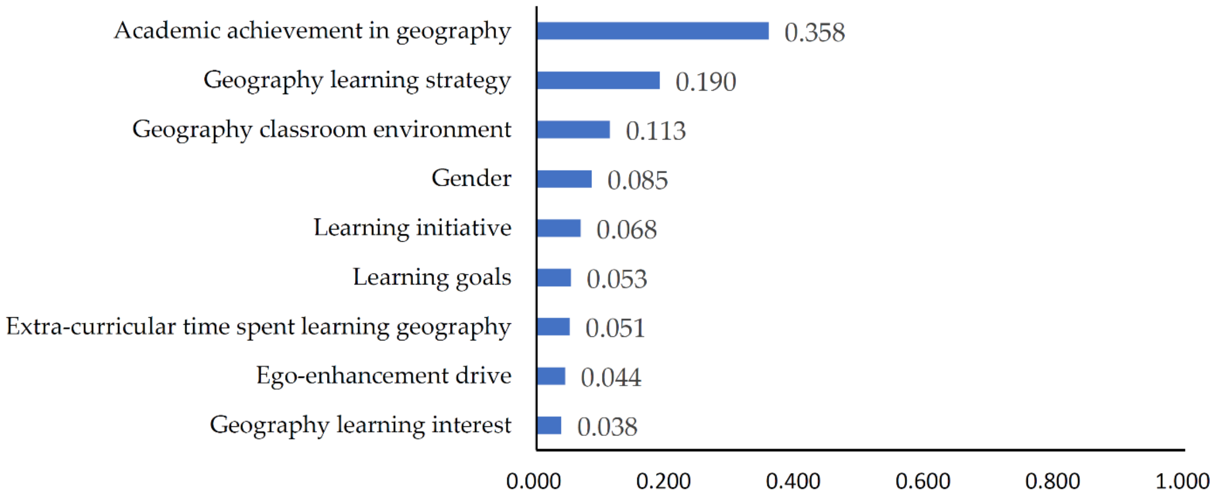

Figure 3. The predictor variables, in order of importance

The Evaluation of the Prediction Model

The confusion matrix and classification accuracy are presented in Table 3 and Table 4, respectively. The model accuracy for the testing data set is 81.25%. According to the definition of precision and recall, the model's precision with the testing data set was 84.49% and the model recall of the testing data set was 94.05%.

Table 3. Confusion matrix

		Predicted class	
		low	**high**
Actual class of training data	low	48	112
	high	31	630
Actual class of testing data	low	11	29
	high	10	158

* Accuracy refers to the proportion of correctly classified cases compared to the total number of cases. Thus, accuracy = (11+158)/(11+29+10+158).
* Precision refers to the proportion of cases with positive prediction that are real positive samples. Thus, precision = 158/(29+158).
* Recall refers to the proportion of positive cases in the sample that are predicted correctly. Thus, recall = 158/(10+158).

Table 4. Classification accuracy

		Number	**Proportion**
Training data	Correct	678	82.58%
	Wrong	143	17.42%
	Total	821	
Testing data	Correct	169	81.25%
	Wrong	39	18.75%
	Total	208	

ROC curve for the decision tree is shown in Figure 4. The value of the AUC was the model's direct output. In this classification, the AUC of the training data set is 0.684 and the AUC of the testing data set is 0.657.

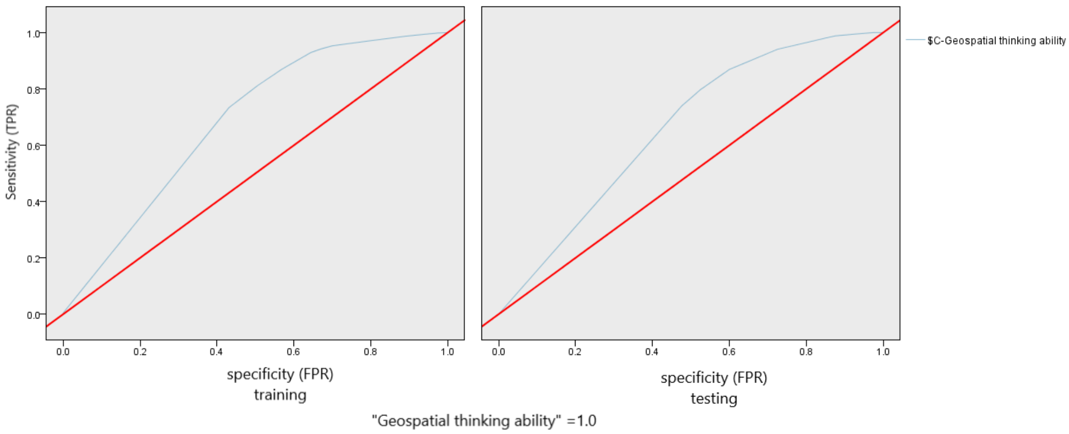

Figure 4. ROC curve for the decision tree.

Discussion

Using a decision tree with a C5.0 algorithm, we constructed a nine-factor model for predicting geospatial thinking ability, and evaluated the contribution of these factors. All findings in this paper are derived from the literature and related data analysis.

First, our results show that the model can effectively predict students' geospatial thinking ability. The accuracy, precision, and recall values of the model are all greater than 80%, showing a good result. According to Fawcett (2006), the AUC value of our model for predicting geospatial thinking ability is 0.657, which is greater than 0.5, indicating that the model is more accurate than a random guessing.

Second, academic achievement in geography is the most important predictor among the nine variables selected by our model. This is consistent with the research by Wan et al. (2017), who found that students' academic performance in the topic of geography has a large and stable predictive effect on their geospatial thinking ability. Given the spatial nature of geography learning, we can understand the close ties between academic achievement in the field of geography and geospatial thinking ability. Learning geography can help students understand the concepts of region, space, and environment, as well as understand the logic of spatial relations (Aliman, Budijanto, Sumarmi, Astina, & Arif, 2019; Metoyer & Bednarz, 2017). Geospatial thinking is developed through the process of using spatial knowledge to solve problems (Gauvain, 1993). Therefore, it is reasonable to believe that students' achievement in geography learning is an important factor when it comes to predicting their geospatial thinking ability.

Third, geography learning strategy is the second most important factor in the prediction model. Students who adopt deep strategies and surface strategies have different ways of learning, resulting in different achievements (Caballos & Esteban, 1988; Jung, 2015; Yip, 2013). Students using a deep learning strategy tend to understand and reflect their knowledge to the maximum extent (Chan et al., 2012). They pay more attention to the logical connection of the occurrence and development of geographical phenomena rather than simply memorizing knowledge, so their geospatial thinking ability tends to be more developed. Therefore, there are circumstances in which students’ learning strategy predicts their geospatial thinking ability.

Fourth, the geography classroom environment is the third most important factor. Previous studies pointed out that an orderly classroom is a necessary condition for effective teaching and student learning (Gaskins, Herres, & Kobak, 2012). Conversely, a chaotic classroom environment affects the teaching quality (Borg, Riding, & Falzon, 1991) and learning outcomes (Infantino & Little, 2005; Pianta, Belsky, Vandergrift, Houts, & Morrison, 2008). Our results support the findings of previous studies (e.g. Borg et al., 1991). The geography classroom is the main place where students are trained in geospatial thinking. When there is order in the classroom and when teachers provide sufficient guidance and timely feedback, students can better learn geospatial thinking skills and correct errors in time. Therefore, the geography classroom environment is relatively important for predicting geospatial thinking ability.

Fifth, cognitive drive was not included as a variable in the prediction model, indicating that cognitive drive has little predictive power for geospatial thinking ability. Theoretically, cognitive drive is the direct internal motivation for students to acquire knowledge and the basic factor for generating learning results (Shi, 1994). Studies have also showed a positive correlation between students' internal motivation and their academic performance (Areepattamannil, Freeman, & Klinger, 2011; Guo & Cao, 2019; Zhang & Shen, 2005; Zhu, Han, Qian, Shi, & Yuan, & 1987). Thus, cognitive drive is an important factor for geospatial thinking ability but it is not an indicative factor.

Limitations and Future Directions

This study has three limitations that need to be addressed. First, this research is a cross-sectional study reflecting the prediction of geospatial thinking ability at a specific time. Second, our sample (n = 1029) was randomly selected from 27 classrooms in different regions of China, which can be generalized statistically. Nevertheless, the robustness of the prediction model needs to be tested in other areas. Third, the factors influencing geospatial thinking ability contained in this research are limited. However, some of the mechanisms influencing geospatial thinking ability are so complex that they may not have been revealed (Ishikawa, 2013; Lee & Bednarz, 2012; Lobben & Lawrence, 2015). Hence, our prediction model needs further improvement. Nevertheless, the present study provides a new path for the prediction of geospatial thinking ability. In the future, we can improve the prediction model through dynamic tracking, by expanding the sample size and diversity, and by incorporating additional predictors.

Conclusion

We established a model with nine factors predicting geospatial thinking ability based on the decision tree with the C5.0 algorithm. The results show that our model is able to predict students' geospatial thinking ability with an accuracy rate of 81.25%. Meanwhile, the model revealed the three most important predictors of geospatial thinking ability, which are academic achievement in geography, geography learning strategy, and geography classroom environment. Moreover, gender, learning initiative, learning goals, extra-curricular time spent learning geography, ego-enhancement drive, and geography learning interest can also predict geospatial thinking ability, but to a lesser extent.

Suggestions for Geography Teachers

Based on the results of our study, we propose a set of suggestions for geography teachers. First, geography teachers should be sensitive to the students' academic achievement and spatial knowledge acquisition in geography. Second, teachers should encourage students to adopt effective learning strategies. Specifically, they can encourage students to think deeply about geographical problems through inquiry, by introducing geospatial technology, and by encouraging them to reflect over geographical principles, and by encouraging the combination of new geographic knowledge with existing knowledge and experience. Third, teachers should be mindful of their classroom environment. We suggest that teachers should maintain order in the geography classroom and give students sufficient and timely guidance, thus creating a good learning environment. Fourth, teachers should pay attention to students' learning motivation, extra-curricular time spent learning geography, and interest in learning geography. If there is a bad situation, teachers should intervene and guide students in a timely manner. It is worth noting that a key factor in improving geospatial thinking ability is to teach students how to use spatial perspective and relevant strategies to solve geographical problems. Lastly, we would advise geography teachers to use the prediction model developed in this paper to predict students' geospatial thinking ability, which can be helpful in preparing and adjusting teaching plans.

Acknowledgements

The research was supported by the Ministry of Education of the People's Republic of China, Humanities and Social Sciences Research Project of the Year 2019 Funding Project (approval number: 19YJA880042) and the "2020 Xing Fu Zhi Hua" Fund Pilot Project (Humanities and Social Sciences) at the East China Normal University (approval number: 2020ecnu-xfzh001). The authors would like to expresses their gratitude to the students who volunteered to participate in this study and to the teachers who helped with this study.

References

Abu, A. (2016). Educational data mining & students' performance prediction. *International Journal of Advanced Computer Science & Applications, 7*(5), 212-220.

Aliman, M., Budijanto, Sumarmi, S., Astina, I. K., & Arif, M. (2019). The effect of earthcomm learning model and spatial thinking ability on geography learning outcomes. *Journal of Baltic Science Education, 18*(3), 323-334.

Anthamatten, P. (2010). Spatial thinking concepts in early grade-level geography standards. *Journal of Geography, 109*(5), 169-180.

Areepattamannil, S., Freeman, J. G., & Klinger, D. A. (2011). Influence of motivation, self-beliefs, and instructional practices on science achievement of adolescents in Canada. *Social Psychology of Education*, *14*(2), 233-259.

Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers and Education, 113*, 177-194.

Atit, K., Uttal, D. H., & Stieff, M. (2020). Situating space: Using a discipline-focused lens to examine spatial thinking skills. *Cognitive Research: Principles and Implications, 5*(1), 1-16.

Ausubel, D. P., Novak, D. J., & Hanesian, H. (1978). *Educational psychology: A cognitive view*. New York: Holt, Rinehart & Winston.

Bednarz, R., & Lee, J. (2019). What improves spatial thinking? Evidence from the spatial thinking abilities test. *International Research in Geographical and Environmental Education, 28*(4), 262-280.

Berberoğlu, G., & Demircioğlu, H. (2000). Factors affecting achievement in general chemistry courses among science major students. *Egitim ve Bilim, 25*(118), 35-42.

Borg, M. G., Riding, R. J., & Falzon, J. M. (1991). Stress in teaching: A study of occupational stress and its determinants, job satisfaction and career commitment among primary schoolteachers. *Educational Psychology, 11*(1), 59-75.

Caballos, A. M., & Esteban, A. (1988). Study skills and problem-solving strategies in Spanish students. *School Psychology International, 9*(2), 147-150.

Catling, S. J. (1978). The child's spatial conception and geographic education. *Journal of Geography, 77*(1), 24-28.

Chan, K. W., Wong, A. K. Y., & Lo, E. S. C. (2012). Relational analysis of intrinsic motivation, achievement goals, learning strategies and academic achievement for Hong Kong secondary students. *Asia-Pacific Education Researcher, 21*(2), 230-243.

Che, D., Liu, Q., Rasheed, K., & Tao, X. (2011). Decision tree and ensemble learning algorithms with their applications in bioinformatics. In H. Arabnia & Q. N. Tran (Eds.), *Software tools and algorithms for biological systems* (pp. 191-199). New York, NY: Springer. doi:10.1007/978-1-4419-7046-6_19

Chen, J., Feng, J., Sun, X., Wu, N., Yang, Z., & Chen, S. (2019). MOOC dropout prediction using a hybrid algorithm based on decision tree and extreme learning machine. *Mathematical Problems in Engineering, 2019*, 1-11. doi:10.1155/2019/8404653

Collins, L. (2018). The impact of paper versus digital map technology on students' spatial thinking skill acquisition. *Journal of Geography, 117*(4), 137-152.

Entwistle, N., McCune, V., & Walker, P. (2001). Conceptions, styles, and approaches within higher education: Analytical abstractions and everyday experience. In R. Stenberg & L. F. Zang, (Ed.), *Perspectives on cognitive, learning and thinking styles* (pp. 211-245). Mahwah, NJ: Lawrence, Erlbaum.

Ergene, T. (2011). The relationships among test anxiety, study habits, achievement, motivation, and academic performance among Turkish high school students. *Egitim ve Bilim, 36*(160), 320-330.

Fawcett, T. (2006) An Introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861-874.

Flynn, K. C. (2018). Improving spatial thinking through experiential-based learning across international higher education setting. *International Journal of Geospatial and Environmental Research, 5*(3).

Francis, B. K., & Babu, S. S. (2019). Predicting academic performance of students using a hybrid data mining approach. *Journal of Medical Systems, 43*(6), 1-15.

Garcia De la Vega, A. (2019). Spatial thinking ability acquisition through geospatial technologies for lifelong learning. In R. De Miguel González, K. Donert, & K. Koutsopoulos (Eds.), *Geospatial technologies in geography education* (pp. 21-40). Cham: Springer Nature Switzerland AG.

Gaskins, C. S., Herres, J., & Kobak, R. (2012). Classroom order and student learning in late elementary school: A multilevel transactional model of achievement trajectories. *Journal of Applied Developmental Psychology, 33*(5), 227-235.

Gauvain, M. (1993). The development of spatial thinking in everyday activity. *Developmental Review, 13*(1), 92-121.

Golledge, R. G. (2002). The nature of geographic knowledge. *Annals of the Association of American Geographers, 92*(1), 1-14.

Guo, K., & Cao, Y. (2019). The influence of learning motivation on students' academic achievement: An empirical study based on large-scale student survey. *Educational Science Research, 30*(3), 62-67. Retrieved from https://www.cnki.com.cn/Article/CJFDTotal-JYKY201903013.htm

Hamoud, A. K., Hashim, A. S., & Awadh, W. A. (2018). Predicting student performance in higher education institutions using decision tree analysis. *International Journal of Interactive Multimedia and Artificial Intelligence, 5*(2), 26-31.

Han, J., Fang, M., Ye, S., Chen, C., Wan, Q., & Qian, X. (2019). Using decision tree to predict response rates of consumer satisfaction, attitude, and loyalty surveys. *Sustainability, 11*(8), 2306. doi:10.3390/su11082306

Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques*. Saint Louis: Elsevier Science & Technology.

Hartshorne, R. (1958). The concept of geography as a science of space, from Kant and Humboldt to Hettner. *Annals of the Association of American Geographers, 48*(2), 97-108.

Hilman, I., & Mainaki, R. (2020). Advantage of map as geography learning media to enhance students spatial intelligence. *International Journal of GEOMATE, 18*(68), 225-232.

Huynh, N. T., & Sharpe, B. (2013). An assessment instrument to measure geospatial thinking expertise. *Journal of Geography, 112*(1), 3-17.

Infantino, J., & Little, E. (2005). Students' perceptions of classroom behaviour problems and the effectiveness of different disciplinary methods. *Educational Psychology, 25*(5), 491-508.

Ishikawa, T. (2013). Geospatial thinking and spatial ability: An empirical examination of knowledge and reasoning in geographical science. *The Professional Geographer, 65*(4), 636-646.

Jung, E. K. (2015). Relation between learning strategy and academic achievement in the dental hygiene students. *Journal of Korean Society of Dental Hygiene, 15*(3), 371-377.

Kali, Y., Orion, N., & Mazor, E. (1997). Software for assisting high-school students in the spatial perception of geological structures. *Journal of Geoscience Education, 45*(1), 10-21.

Kim, M., & Bednarz, R. (2013). Development of critical spatial thinking through GIS learning. *Journal of Geography in Higher Education, 37*(3), 350-366.

Lee, J., & Bednarz, R. (2009). Effect of GIS learning on spatial thinking. *Journal of Geography in Higher Education, 33*(2), 183-198.

Lee, J., & Bednarz, R. (2012). Components of spatial thinking: Evidence from a spatial thinking ability test. *Journal of Geography, 111*(1), 15-26.

Lin, C., Yang, Z., & Huang, X. (2004). *The comprehensive dictionary of psychology*. Shanghai: Shanghai Education Publishing House.

Lobben, A., & Lawrence, M. (2015). Synthesized model of geospatial thinking. *The Professional Geographer, 67*(3), 307-318.

Lozano, A. B., Uzquiano, M. P., Riobo, A. P., Malmierca, J. L. M., & Blanco, J. C. B. (2011). Academic goals of high and low academic achievers in mandatory secondary education and optional advanced secondary education. *Revista de Educación (Madrid), 354*(354), 341-368.

Metoyer, S., & Bednarz, R. (2017). Spatial thinking assists geographic thinking: Evidence from a study exploring the effects of geospatial technology. *Journal of Geography, 116*(1), 20-33.

Miguéis, V. L., Freitas, A., Garcia, P. J. V., & Silva, A. (2018). Early segmentation of students according to their academic performance: A predictive modelling approach. *Decision Support Systems, 115*, 36-51.

Mitchell, M. T. (1997). *Machine learning*. New York: The McGraw-Hill Companies.

Montello, D. R., Grossner, K. E., & Janelle, D. G. (2014). Concepts for spatial learning and education: An introduction. In D. R. Montello, K. E. Grossner, & D. G. Janelle (Eds.), *Space in mind: Concepts for spatial learning and education* (pp. 3-29). Cambridge: The MIT Press.

National Research Council. (2006). *Learning to think spatially*. Washington, DC: The National Academies Press.

Organisation for Economic Co-operation and Development. (2015). Student common part questionnaire of program for international student assessment. Retrieved from https://tilssc.naer.edu.tw/upload/pisa/Chinese_(Taiwan)_For_Student_Questionnaire_Common_Part.pdf

Painho, M., Santos, M. Y., & Pundt, H. (2010). *Geospatial thinking*. Berlin: Springer.

Pianta, R. C., Belsky, J., Vandergrift, N., Houts, R., & Morrison, F. J. (2008). Classroom effects on children's achievement trajectories in elementary school. *American Educational Research Journal, 45*(2), 365-397.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning, 1*(1), 81-106.

Quinlan, J. R. (1992). *C4.5 programs for machine learning*. San Mateo, CA: Morgan Kaufmann.

Quinlan, J. R. (1996). Improved use of continuous attributes in C4.5. *The Journal of Artificial Intelligence Research, 4*, 77-90.

Quinlan. J. R. (2019, April). C5.0: An informal tutorial. Retrieved from https://www.rulequest.com/see5-unix.html

Ramaswami, G., Susnjak, T., Mathrani, A., Lim, J., & Garcia, P. (2019). Using educational data mining techniques to increase the prediction accuracy of student academic performance. *Information and Learning Sciences, 120*(7/8), 451-467.

Shi, L. (1994). *Learning theory: Theories and principles of learning psychology*. Beijing: People's Education Press. Retrieved from https://book.douban.com/subject/3455680/

Shin, E., Milson, A. J., & Smith, T. J. (2016). Future teachers' spatial thinking skills and attitudes. *Journal of Geography, 115*(4), 139-146.

Stieff, M., Origenes, A., DeSutter, D., Lira, M., Banevicius, L., Tabang, D., & Cabel, G. (2018). Operational constraints on the mental rotation of STEM representations. *Journal of Educational Psychology, 110*(8), 1160-1174.

Suguna, R., Shyamala Devi, M., Bagate, R. A., & Joshi, A. S. (2019). Assessment of feature selection for student academic performance through machine learning classification. *Journal of Statistics and Management Systems: Swarm Intelligence & Evolutionary Computation for Problem Solving, 22*(4), 729-739.

Tan, P. N., Steinbach, M., & Kumar, V. (2016). *Introduction to data mining*. New Delhi: Pearson Education India.

Tomaszewski, B., Vodacek, A., Parody, R., & Holt, N. (2015). Spatial thinking ability assessment in Rwandan secondary schools: Baseline results. *Journal of Geography, 114*(2), 39-48.

Uttal, D. H., Miller, D. I., & Newcombe, N. S. (2013). Exploring and enhancing spatial thinking: Links to achievement in science, technology, engineering, and mathematics?. *Current Directions in Psychological Science, 22*(5), 367-373.

Verma, K. (2015). Influence of academic variables on geospatial skills of undergraduate students: An exploratory study. *The Geographical Bulletin, 56*(1), 41-55.

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for stem domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology, 101*(4), 817-835.

Wakabayashi, Y. (2015). Measurement of geospatial thinking abilities and the factors affecting them. *Geographical Reports of Tokyo Metropolitan University, 50*, 127-136.

Wan, J., Lu, X., Lu, Y., Du, F., Wang, J., & Ju, B. (2017). Influencing factors of middle school students' spatial thinking ability: A case study on senior one students of Baiyin No. 1 Middle School in Gansu Province. *Progress in Geography, 36*(7), 853-863. doi:10.18306/dlkxjz.2017.07.007

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques*. Burlington: Morgan Kaufmann Publishers.

Xiong, P. (2011). *Data mining algorithms and clementine practice*. Beijing: Tsinghua University Press. Retrieved from https://book.douban.com/subject/6113968/

Xu, X., Wang, J., Peng, H., & Wu, R. (2019). Prediction of academic performance associated with internet usage behaviors using machine learning algorithms. *Computers in Human Behavior, 98*, 166-173.

Yani, A., Mulyadi, A., & Ruhimat, M. (2018). Contextualization of spatial intelligence: Correlation between spatial intelligence, spatial ability, and geography skills. *Journal of Baltic Science Education, 17*(4), 564-575.

Yip, M. C. W. (2013). Learning strategies and their relationships to academic performance of high school students in Hong Kong. *Educational Psychology, 33*(7), 817-827.

Zhang, H., & Shen, L. (2005). Impacts of motivation and metacognition on study performance. *Journal of Psychological Science, 28*(1), 114-116. doi:10.16719/j.cnki.1671-6981.2005.01.029

Zheng, R. (1994). *Psychological diagnosis of middle school students*. Jinan: Shandong Education Press. Retrieved from https://book.douban.com/subject/1074423/

Zhu, Z., Han, F., Qian, Q., Shi, T., & Yuan, Y. (1987). Research on the quantification of students' cognitive drive. *Journal of Psychological Science, 10*(6), 54-56. Retrieved from https://www.cnki.com.cn/Article/CJFDTOTAL-XLKX198706013.htm