Research on User Profile Analysis Method Based on LGIM Model

Teng Zong[1,2](✉), Fengsi Wang[3], Liang Guo[1], Yibo Liu[1], Xiaonan Feng[2], Zifu Qin[2], and Yinxiang Xia[4]

[1] National University of Defense Technology, Wuhan 430000, China
zt@nudt.edu.cn
[2] Information Engineering University, Zhengzhou 450000, China
[3] Henan University, Zhengzhou 450000, China
[4] Zhejiang Provincial Military Command, Hangzhou 310000, China

Abstract. Student profile analysis is a depiction of the characteristics of individual learning behavior and group learning development law of students during college. This study focused on 12,181 students surveyed between 2018 and 2021, and mainly based on the LGIM model. We tried to conduct in-depth research and analyze the characteristics of student profiles from multiple dimensions. Firstly, LGIM Model is tested for moderating effects based on categorical variables, and the models of management and engineering students are compared and analyzed. Then, the trend of students' learning gains was analyzed for different statistical years and student grades. Through the above analysis, this study scientifically summarizes the learning characteristics of college students in different disciplines during their studies, the development characteristics and transformation direction of school education concepts, and the general laws of students' learning and growth during college. The research results are helpful to improve the exploration and utilization of students' potential, improve the quality of university education, and provide some reference for teachers to teach according to their aptitude.

Keywords: User profile · Student characteristics · Moderating effects · LGIM model

1 Introduction

In the era of big data, information and knowledge play an increasingly important role in developing economies and societies. Effective access to and utilization of information resources has become the most important embodiment and guarantee for the development of individuals and organizations in this era. The value of information is reflected in all aspects of our lives and learning. According to the China Digital Economy Development Research Report (2023), the domestic digital economy continues to grow rapidly, and the scale of the digital economy exceeds 50 trillion yuan for the first time, accounting for 41.5% of GDP [1]. In the field of education, Ministry of Education of China formulated and released the Digital Literacy of Teachers in 2022 [2]. All kinds of education data also play an important role in effectively promoting the national education

Y. Zhang et al. (Eds.): CENet 2023, LNEE 1125, pp. 52–62, 2024.
https://doi.org/10.1007/978-981-99-9239-3_5

digitalization strategic action, improving the education informatization standard system, and enhancing teachers' awareness, ability and responsibility to use digital technology to optimize, innovate and change education and teaching activities.

Data is the foundation of all information services. As big data enters a period of accelerated development around the world, data-based user services are getting closer to the personalized needs of users. Research on user profiling is attracting more and more researchers' attention. As a tool to achieve accurate information services, user profiles have been widely used in many fields in recent years [3]. Relevant studies have shown that constructing user profile models can help better understand user needs and achieve personalized and accurate information services.

The essence of user profile research is to extract the characteristics of users and classify users. Explore users' personalized needs based on their various behaviors, habits and preferences. The user profile analysis of student characteristics is essentially based on the potential characteristics of the student group to achieve scientific classification of the student group. The student profile analysis provides a basis for teachers to teach according to aptitude, and can also be used as a means of educational evaluation to test the characteristics of school development. At the same time, student profiles reflect the development trend and general law of students in the whole university stage. Through the profound analysis, this study provides a basis for students, teachers and education administrators to improve learning gains, and has important practical significance for giving full play to students' interests and characteristics, creating a good learning atmosphere, improving the quality of talent training, and promoting the benign development of university education.

2 Research

2.1 Review of Related Researches

User profiles were used in the field of product design and marketing in the early stage. Through user research, questionnaire interviews and other methods to explore user demands, outline target user profiles, so that product design does not depart from user and market demand, and then help enterprises achieve refined operation and marketing. With the emergence of various data mining technologies, it has brought new vitality to user profile research. In the big data environment, researchers analyze users' basic attributes, social attributes, behavioral habits, interests and hobbies from massive user behavior data through data mining and analysis methods, and refine user personalized labels to build more accurate user profiles. At the same time, the application field of user profiles is constantly expanding, from e-commerce, social networks to teaching practices [4].

From the perspective of user profile analysis technology, Zhuang Zhang (2020) proposed the idea of cross-modal learning and designed a user profile model based on multimodal fusion in view of the problem that the modal information in user profile work cannot be fully utilized [5]. Zhang (2020) aims at the problem of imperfect user preference acquisition in user-based collaborative filtering algorithms, and proposes a KNN classification recommendation algorithm based on dynamic user profile labels to solve the problems of ignoring users' potential preferences and changing trends of user preferences in current mainstream recommendation algorithms [6]. Jiang (2016) studies the information ontology extraction method based on user profile by constructing a mathematical model of behavior-theme-vocabulary trinity, constructing user profiles, and realizing intelligent information push [7].

From the perspective of user profiles application, Lin (2018) aimed at social media applications, takes Weibo as an example to explore topics that users are interested in by collecting and analyzing the dynamics of users' Weibo, and to build Weibo user portraits, which plays a certain role in the personalized information service of social media and public opinion governance [8]. Wei (2021) proposed an accurate recommendation method of remote sensing information based on user profile for the accurate service of remote sensing information, which uses the theme model to construct the user profile by collecting and analyzing the explicit and implicit feedback behavior of the user, and completes the accurate recommendation of remote sensing information according to the profile model [9]. In the existing research, the application of user profiles in teaching practice is relatively rare. However, teaching according to aptitude has been an ideal teaching method in the field of education since ancient times, so this paper chooses this perspective as the research direction.

2.2 Research Data

The data used in this study are collected between 2018 and 2021 in the Chinese College Student Survey (CCSS) of X University. CCSS is an authoritative survey scale for studying the learning and development of college students in China, which was developed by Tsinghua University on the basis of the internationally influential National Survey of Student Engagement (NSSE) [10]. X University surveyed 12,181 students over four years. According to the discipline types, 7736 engineering students and 4445 management students were obtained. In the four years, 3566 students were surveyed in 2018, 2309 students in 2019, 3275 students in 2020, and 3031 students in 2021. Among all respondents, a total of 3588 first-graders participated in the survey, 3366 second-graders, 2883 third-graders and 2344 fourth-graders. Through the above classification, it lays a foundation for the subsequent potential profile analysis based on different student characteristics, which is convenient for studying the educational development law under different student characteristics.

3 Method

3.1 LGIM Model

On the basis of educational evaluation research over the years, Zong (2023) proposed Quality of Student Involvement (QSI), Quality of Teacher Involvement (QTI), and Supportive Campus Environment (SCE) has become the three key factors for universities to improve the quality of talent training, and is the main influence factor of Student Learning Gains (SLG). From the perspective of teachers, QTI is divided into Effective Teaching Practices (ETP) and Emotional Support of Teachers and Students (EST-S) [11]. Through data cleaning and confirmatory factor analysis, the evaluation indexes of student learning and development with good reliability and validity are obtained. On this basis, the Learning Gains Influence Mechanism Model (LGIM Model) was explored by stepwise hypothesis testing (see Fig. 1). Through verification, the model fit indicators perform well. In the full sample analysis, the interpretable variance of LGIM Model reached 84%, which was significantly better than other models in similar studies [12]. Subsequent profiling of student characteristics is carried out on the basis of LGIM Model.

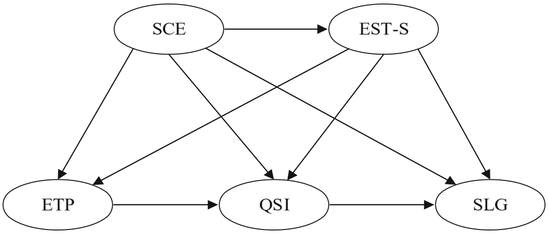

Fig. 1. Learning gains influence mechanism model (LGIM model).

3.2 The Basic Theory of Moderating Effect

As shown in Fig. 2, in model analysis, if the relationship between two variables (such as the relationship between X and Y) is a function of another variable M. In other words, the influence relationship between independent variable X and dependent variable Y is affected by another variable M, so we call M a moderating variable, and this effect is a moderating effect [13]. Among them, the moderating variable M can be categorical (e.g., gender, race, class, etc.) or continuous (e.g., height, age, years of education, etc.), which affects the direction (positive or negative) and strength of the relationship between the independent variable X and the dependent variable Y. The independent variable X can also be categorical or continuous; However, the dependent variable Y can only be continuous.

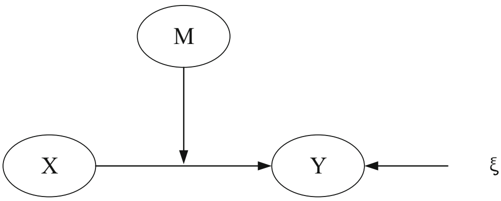

Fig. 2. Schematic diagram of moderating effect.

Baron et al. mentioned that the moderating effect affects the slope of the relationship between the independent variable X and the dependent variable Y [14]. In particular, Y and X have the following relationship:

$$Y = \beta_0 + \beta_1 X + \beta_2 M + \beta_3 MX + e \tag{1}$$

As shown in Fig. 3, the path analysis diagram of the moderating effect is shown. Equivalently varying the above formula yields:

$$Y = (\beta_0 + \beta_2 M) + (\beta_1 + \beta_3 M)X + e \tag{2}$$

We can get that when M is a fixed value, this is a linear regression of Y to X. In general, the relationship between Y and X is characterized by the regression coefficient $(\beta_1 + \beta_3 M)$. The regression coefficient is a linear function of M. If not 0, M is the regulating variable, which reflects the size of the regulating variable.

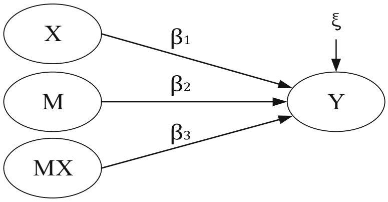

Fig. 3. Path analysis diagram of moderating effects.

3.3 Tests for Moderating Effects

Usually, different characteristics of students tend to show different atmosphere. Through group comparison of students, the characteristics of LGIM Model presented in different situations are discussed, and the development rules of students are summarized. In college, the major of the student is the main difference. Firstly, it is tested whether different disciplines have a moderating effect on the influence relationship of each path coefficient in the model. In this study, management students and engineering students were selected for comparative testing. According to the path order of LGIM Model, nine comparison models are constructed. From 1 to 9, it represents the difference comparison of path coefficients between the two variables. From this, the null hypothesis $H_0 1$ to $H_0 9$ is established (see Table 1). Taking $H_0 1$ as an example, it means that the discipline type makes the impact factor between SCE and EST-S significantly different. The rest of the assumptions correspond to this.

P value is the probability of the null hypothesis misestimation. Usually, $P > 0.05$ indicates the rejection of the null hypothesis, and $P \leq 0.05$ indicates acceptance of the null hypothesis [15]. According to this judgment basis, we get that under different disciplines, except for the influence relationship of EST-S on QSI, the influence relationship of SCE on QSI and the influence relationship of ETP on QSI, the other influence relationships have significant differences. In other words, differences disciplines have significant differences on QTI and SLG, except that they have no significant differences on QSI.

Table 1. Test results of moderating effect.

Assumptions of comparison models	Representative path	Chi-square value (χ^2)	*P*-value
$H_0 1$	SCE → EST-S	26.115	0.000
$H_0 2$	EST-S → ETP	35.606	0.000
$H_0 3$	SCE → ETP	18.669	0.000
$H_0 4$	EST-S → QSI	0.032	0.857
$H_0 5$	SCE → QSI	0.320	0.571
$H_0 6$	EST-S → SLG	9.086	0.003
$H_0 7$	SCE → SLG	16.169	0.000
$H_0 8$	ETP → QSI	2.752	0.097
$H_0 9$	QSI → SLG	33.006	0.000

Note $N_{\_Management} = 4445$, $N_{\_Engineering} = 7736$

4 Results

4.1 Student Profiling for Different Disciplines

Through testing, we find that different disciplines have a moderating effect on the learning gains influence mechanism. The path represented by the bold lines in Fig. 4 is the path with the moderating effect verified above. This shows that management students and engineering students show significant differences in the process of learning gains. The analysis is mainly based on the following two points:

First, for the dependent variable SLG, the path coefficients of the three direct impact factors (SCE, EST-S and QSI) of management students were 0.31, 0.35 and 0.36, respectively, and the path coefficients of the three direct impact factors (SCE, EST-S and QSI) of engineering students were 0.24, 0.28 and 0.49, respectively. In contrast, it can be seen that the influence of engineering students QSI on SLG is significantly higher than that of the other two variables. The influence of management students on SLG in QSI, EST-S and SCE is basically the same.

Second, the influence relationship between the three factors of SCE, EST-S and ETP is a typical mediation model with moderating effect. Wherein, the independent variable is SCE, the dependent variable is ETP, the mediator variable is EST-S, and the moderator variable is discipline type. Among them, the moderator variable acts on both direct and indirect effects, which makes management students and engineering students show different characteristics. The direct impact effect of engineering students is more significant, while the indirect effect of management students is more significant. The most significant difference in impact is reflected in the path of EST-S on ETP. The path coefficient for management students is 0.35 and 0.19 for engineering students. In contrast, management students have a significantly stronger impact than engineering students.

After the above analysis, it shows that engineering students are more inclined to obtain learning gains through personal efforts to improve the quality of their learning input. Management students have relatively little personal investment, but they can make more effective use of the advantages of school resources and are better at handing the relationship between teachers and students. Moreover, the establishment of a good teacher-student relationship is more likely to affect the change of teaching behavior, so that teachers can make appropriate adjustments in a more timely manner in the classroom teaching process and improve the classroom teaching input.

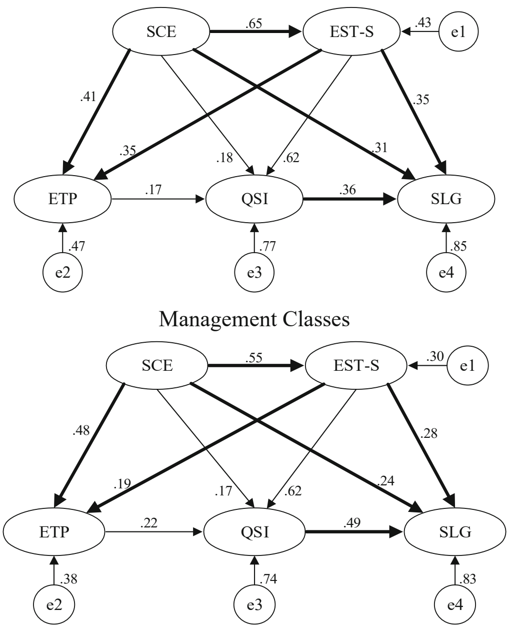

Fig. 4. Comparative analysis of LGIM Models for different disciplines.

4.2 Student Profiling for Different Statistical Years

Based on the CCSS data from 2018 to 2021, models of different statistical years can be compared. Analysis differed from follow-up studies for different statistical years. Because the follow-up study is a fixed sample, the sample measured in this study each year is a different group of students. The four-year data are stratified according to the discipline type, and the results can more generally reflect the overall change law.

Firstly, the data from 2018 to 2021 were tested sequentially by LGIM Model, and the explainable variances of SLG were 83%, 87%, 83%, and 82%, respectively. The difference of interpretability between statistical years is relatively small, and longitudinal comparison of explanatory proportion of each variable can be made for four consecutive years.For visual purposes, this study converted the path coefficients of influence variables in different statistical years into corresponding influence proportions, as shown in Table 2.

Table 2. Analysis of SLG impact in different statistical years.

The variable name	2018		2019		2020		2021	
	Path factor	Impact ratio (%)	Path factor	Impact ratio (%)	Path factor	Impact ratio (%)	Path factor	Impact ratio (%)
QSI	0.484	48.0	0.447	43.7	0.463	45.7	0.405	39.3
EST-S	0.286	28.4	0.271	26.5	0.286	28.2	0.352	34.1
SCE	0.238	23.6	0.306	29.9	0.264	26.1	0.274	26.6

Note $N\_{2018} = 3566$, $N\_{2019} = 2309$, $N\_{2020} = 3275$, $N\_{2021} = 3031$

The impact ratio of QSI, EST-S and SCE on SLG over four years is plotted as a line graph (see Fig. 5). We conclude that QSI has always been the variable with the greatest influence on SLG during the four years, but the explanation proportion of QSI decreases, the explanation proportion of EST-S increases, and the explanation proportion of SCE tends to be stable.

The above analysis shows that in recent years, with the change of educational philosophy from "teaching" to "learning", teachers are paying more attention to emotional communication with students. In the process of learning gains, students gradually change from "independent learning" which is dominated by self-learning to "compound learning" which synchronously develops with self-learning and interactive learning. Students have a greater sense of participation and acquisition in the learning process, rather than just passively accepting knowledge.

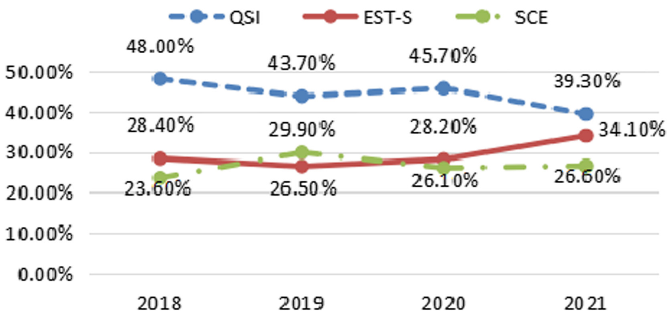

Fig. 5. Analysis of the trend of SLG influence in different statistical years.

4.3 Student Profiling for Different Grades

The student profiling for different grades reflects the law of students' growth and development in four years. Firstly, LGIM model was used for the sequence test of data from freshmen to seniors. The explainable variances of SLG were 85%, 84%, 83% and 84%, respectively. It is also concluded that there is no significant difference in the interpretability of SLG among different grades, and horizontal comparison can be made on the interpretability ratio of variables in the four grades. The path coefficients of each influence variable of SLG in different grades were converted into corresponding influence proportions, as shown in Table 3.

Table 3. SLG impact analysis by grade level.

The variable name	Freshman year		Sophomore year		Junior year		Senior year	
	Path factor	Impact ratio (%)	Path factor	Impact ratio (%)	Path factor	Impact ratio (%)	Path factor	Impact ratio (%)
QSI	0.461	45.2	0.492	48.6	0.413	40.6	0.472	46.7
EST-S	0.312	30.6	0.264	26.1	0.302	29.7	0.251	24.9
SCE	0.246	24.1	0.256	25.3	0.302	29.7	0.287	28.4

Note $N\_{Freshman} = 3588$, $N\_{Sophomore} = 3366$, $N\_{Junior} = 2883$, $N\_{Senior} = 2344$

In different grades, the impact ratio of QSI, EST-S and SCE on SLG is plotted as a line graph (see Fig. 6). We conclude that QSI has always been the most important direct impact on SLG in the whole college experience. From freshman to sophomore, the influence of QSI on SLG showed an upward trend and reached the highest value in the four years of college. From the sophomore year to the junior year, the influence strength continued to decline, reaching the lowest value in the whole university. From the junior to the senior year, the influence strength picked up again and reached the second peak. In addition, the explanatory strength of EST-S was significantly higher than that of SCE in the freshman year, but with the transition from freshman to senior year, the explanatory strength of SCE gradually improved and surpassed EST-S.

The above analysis shows that from freshman to sophomore year is the "adaptation period" for students to university life. From sophomore to junior year, students have a "slack period" of personal investment. At this stage, because many students have neither the enthusiasm for learning when they first enter college nor the pressure to graduate, the polarization is serious. On the whole, QSI has significantly declined in the interpretation of SLG. In the senior year, as graduation approaches, academic pressure increases, and QSI also increases significantly.

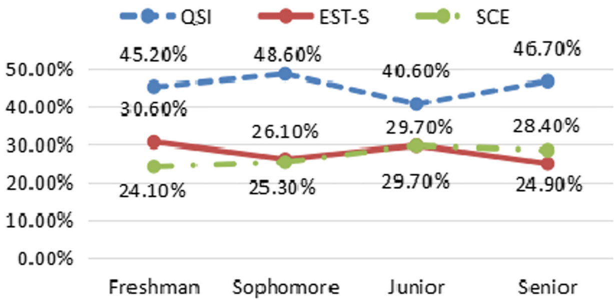

Fig. 6. Analysis of SLG influence trend in different grades.

5 Conclusion

This study mainly carried out a further study on the basis of LGIM Model, and analyzed the characteristics of student profiles from multiple dimensions. After testing the moderating effect of LGIM Model, we conclude that discipline type is a moderator variable of the learning gains influence mechanism. By comparing the study subjects in multiple categories, the following main conclusions were obtained. From the perspective of different disciplines, engineering students are more dependent on learning gains by improving the quality of individual learning input. Management students not only rely on individual efforts, but also focus on the support of the school environment and the emotional interaction between teachers and students. From the perspective of different statistical years, with the transformation of teaching philosophy from "teaching" to "learning", teachers pay more attention to emotional communication with students. When students acquire learning gains, they gradually change from "independent learning" which is dominated by self-learning to "compound learning" which synchronously develops with self-learning and interactive learning. From the perspective of different grades, students have a "slack period" in personal input during the transition from sophomore to junior year. At this stage, the student's own personal input has declined markedly in the interpretation of learning gains, reaching the lowest point throughout the university period. Through the analysis of the above student profiles, this study explains the learning characteristics of college students in different disciplines, the development directions of school education, and the general laws of students' learning and growth in college.

References

1. China Digital Economy Development Research Report: http://www.caict.ac.cn/kxyj/qwfb/bps/202304/t20230427_419051.htm (2023). Last accessed 25 May 2023
2. Notice of the Ministry of Education on the release of Standards for the education industry called Digital Literacy of Teacher: http://www.moe.gov.cn/srcsite/A16/s3342/202302/t20230214_1044634.html. Last accessed 25 May 2023
3. Xu, F., Ying, J.R.: Overview of user profile research at home and abroad. Res. Libr. Sci. **479**(12), 7–16 (2020)

4. Renwu, W., Wenhui, Z.: Behavior and interest labeling construction and application of academic user profiles. J. Mod. Inf. **39**(09), 54–63 (2019)
5. Zhuang, Z., Xiaonian, F., et al.: User profiling based on multimodal fusion technology. Acta Scientiarum Naturalium Universitatis Pekinensis **56**(01), 105–111 (2020)
6. Zhang, Y., Zhang, C., et al.: KNN product recommendation algorithm based on dynamic user profile label. Inf. Sci. **38**(08), 11–15 (2020)
7. Jiang, J., Li, J., et al.: Intelligent information push method based on user profile. Cyber Secur. Data Gov. **35**(23), 86–89+92 (2016)
8. Lin, Y., et al.: User profile of diversified groups in micro-blog based on social identity theory. Inf. Stud. Theory Appl. **41**(03), 142–148 (2018)
9. Wei, E., Liu, D., et al.: Design of precise service system based on user profile for remote sensing information. Radio Eng. **51**(8), 720–724 (2021)
10. Construction of research project on Learning and Development of Chinese University Students: https://www.ioe.tsinghua.edu.cn/info/1141/1046.htm. Last accessed 26 May 2023
11. Zong, T., Wang, F., et al.: Mechanisms influencing learning gains under information security: structural equation modeling with mediating effect. Intell. Autom. Soft Comput. (6), 3447–3468 (2023)
12. Kwan, J.L.Y., Chan, W.: Comparing squared multiple correlation coefficients using structural equation modeling. Struct. Equation Model. **21**(2), 225–238 (2014)
13. Wen, Z., Liu, H: Analyses of Mediating and Moderating Effects Methods and Applications, 2nd edn. Educational Science Publishing House, Beijing (2020)
14. Hayes, A. F.: Beyond Baron and Kenny: statistical mediation analysis in the new millennium. Commun. Monogr. **76**(4), 408–420 (2009)
15. Bonovas, S., Piovani, D: On P-values and statistical significance. J. Clin. Med. **12**(3), 900 (2023)