BITNET
Jurnal Pendidikan Teknologi Informasi
Vol. 10 No. 2, June 2025, pp.12-22
https://journal.umpr.ac.id/index.php/bitnet/issue/...
DOI: https://doi.org/...

The Impact of Basic English Proficiency on the Academic Performance of Information Technology Students

Jong Jek Siang[1]*, Friska F Nainggolan[1], Arida Susyetina[2]
[1*,1] Universitas Kristen Duta Wacana, Fakultas Teknologi Informasi, Program Studi Sistem Informasi
[2] Universitas Kristen Duta Wacana, Fakultas Kependidikan dan Humaniora, Program Studi Pendidikan Bahasa Inggris
* Corresponding Author: jjsiang@staff.ukdw.ac.id

ABSTRACT

The Faculty of Information Technology at a private university in Yogyakarta offers two study programs: Informatics and Information Systems. To enhance students' English proficiency, all first-semester students are required to participate in the Introduction to College English (ICE) program, which comprises levels 1, 2, 3, before taking the English for Specific Purposes (ESP) course. A placement test is administered before the start of the first semester to determine students' initial English proficiency level. This study examines the relationship between students' basic English skills—measured by their ICE placement level—and their academic performance, represented by their first-semester Grade Point Average (IPS). An ANOVA test was conducted on a dataset of 998 students from 2017 to 2021 to determine the significance of this relationship, followed by a Pairwise t-test to identify specific proficiency levels that contributed to significant differences in IPS scores. Data preprocessing steps included cleaning, integration, reduction, and transformation. The ANOVA results revealed a strong correlation between English proficiency and academic achievement, with a p-value (PR > F) of 0%. The Pairwise t-test results indicated that the most significant differences in IPS scores were observed between students with low English proficiency (level 1) and those with higher proficiency (level 3 or ESP). These findings highlight the critical role of English language skills in the academic success of Information Technology students.

Keywords: ANOVA, t-Test Pairwise, ICE, Information Technology

Article history

Received:	*Revised:*	*Accepted:*	*Published:*
2 February 2024	*11 March 2024*	*15 April 2025*	*2 June 2025*

INTRODUCTION

The quality of education is very important for improving the quality of a nation (Muhardi, 2004). There are various factors that can affect students' academic performance. One of them is English language skills. Communication skills in English are not only a necessity but also a necessary competency in today's era of globalization. Therefore, English language development programs are considered important as a foundation to support students in lectures in each study program. (Fenton-Smith, Humphreys, & Walkinshaw, 2015).

The Faculty of Information Technology (FTI), which offers the Informatics and Information Systems programs, requires all incoming students to complete the Introduction to College English (ICE) program—comprising Levels 1, 2, and 3—before advancing to the English for Specific Purposes (ESP) course.

Before starting their first semester, all new students must take the ICE placement test, conducted during each new student admissions (PMB) period. This test, similar to the TOEFL,

assesses students' basic English proficiency through three sections: listening, reading, and language use. The ICE level is determined based on the number of correctly answered questions out of 50: Level 1 (score 1-15), Level 2 (score 16-30), Level 3 (score 31-45), and ESP (score 46-50). Students must reach the ESP level to enroll in compulsory English courses at FTI. Those who do not achieve this level are required to complete the ICE program according to their assigned level. Each ICE level carries 0 credits and is conducted over one full semester.

The results of the ICE placement test are collected and managed by the Language Training Center (LTC), the unit responsible for ICE data management. LTC compiles a summary of scores for each student, including their ICE level based on test performance. These ICE records are stored in a database alongside other academic information managed by LTC.

The university has a Campus Information Service and Intranet Center Unit (PUSPINdIKA) that manages student academic data. Structurally, the PUSPINdIKA unit is separate from the LTC. ICE placement test data is stored by LTC while student academic score data is stored by PUSPINdIKA in 2 (two) different databases. In this study, each data was taken separately in MS Excel format..

Research on the relationship between basic English language skills and academic achievement has been conducted in various countries, such as Thailand (Rudd & Honkiss, 2020), Malaysia (Suyansah & Gabda, 2020) as well as Indonesia (Desiani, Yahdin, & Rodiah, 2020). The results of this study consistently show that basic English language skills affect students' academic achievement. Neumann's research in Canada also found that students' GPA depended on their English language skills (Neumann, Padden, & Mcdonough, 2019) .

The ANOVA method has been used in education and various other fields, including comparing students' perceptions of academic guidance (Sherlywati, Junita, Kristine, & Wisamtamma, 2021), analyzing student learning outcomes (Sihotang, 2022), analyzing the application of learning methods and self-confidence towards English learning achievement (Iswindarti, Rufii, & Hartono, 2020), role playing method for English speaking ability (Sunandar, Alvarez, & Cardozo, 2024)and also evaluating the relationship between procrastination and burnout (Matthew & Widjaja, 2022)

In this study, an analysis was conducted to test whether there is a positive relationship between basic English language skills and their achievement index in the first semester. The test was conducted using the ANOVA method (Albright & Winston, 2019) and Pairwise t-test (Berenson , Levine , & Stephan, 2020). This study also evaluates whether similar results found in different countries also apply to the university. ICE results were chosen as the basic English proficiency variable because the ICE test was conducted before students started their studies. First semester GPA (IPS 1) was chosen as the academic achievement variable because GPA in subsequent semesters may be influenced by other variables besides basic English proficiency.

RESEARCH METODOLOGY

Data Collection

The data obtained in this study were FTI students in 2017-2021. Given the different storage databases, the data obtained were in the form of 2 different Excel files, each of which was obtained from 2 different units, namely the PUSPINdIKA Unit and the Language Training Center (PPB) Unit. Data from the PUSPINdIKA Unit is student academic data including student numbers (NIM), study program, student status in semesters 1 and 2 (Status_1, Status_2), IP semester 1 dan 2 (IPS_1, IPS_2), IP kumulatif 1 dan 2 (IPK_1, IPK_2), and the number of credits for semesters

1 and 2 (Jum_sks_1, Jum_sks_2). The data fields obtained from the LTC are registration number, NIM, study program, class, ICE Placement Test (IC) score, and ICE level (level).

Before the data is further processed, a data preprocessing stage is required which aims to prepare the data so that it is ready to be analyzed. The data preprocessing stage consists of data cleaning, data integration, data reduction, and data transformation. All data processing is done with the help of Python which can perform all stages of preprocessing, data analysis with ANOVA, paired t-test, and create appropriate graphs..

Data Cleaning

At this stage, data checking and cleaning are carried out by deleting data that has missing values, NaN values and data with a value of 0 on cumulative IP 1 and ICE Placement Test value. The value "0" on cumulative IP 1 is assumed that the student has dropped out of college so that the student data is not valid for use. The following is the source code to delete NaN data and "0" value data on the cumulative IP 1 (IPK_1) attribute and ICE Placement Test (IC) value.

Program 1. Data Cleaning

```
akm_ice_df.isnull().sum()
drop_df = akm_ice_df.dropna(subset=('IPK_1', 'IC', 'level'),axis = 0)
for x in drop_df.index:
  if drop_df.loc[x, "IPK_1"] == 0:
    drop_df.drop(x, inplace=True)
drop_df[drop_df['IPK_1'] == 0]
for x in drop_df.index:
  if drop_df.loc[x, "IC"] == 0:
    drop_df.drop(x, inplace=True)
drop_df[drop_df['IC'] == 0]
```

Data Integration

Data integration is done by combining 2 xxcel academic data from the PUSPINdIKA Unit and placement test data from LTC into one integrated dataset. The goal is to avoid redundancy and inconsistency in the dataset to be processed. (Han, Kamber, & Pei, 2011). This process also identifies and eliminates data duplication. In academic data and ICE data, the same NIM and study program attributes are stored, which can cause redundancy in the dataset.

Program 2. Data Integration

```
akm_ice = pd.merge(akm_df,ice_df,on=["NIM","Program_Studi"],how='inner',
indicator=True)
```

Data Reduction

In this process, data sorting is carried out by removing irrelevant or unnecessary data to reduce the number of attributes in the dataset to make it more concise. Because the analysis was carried out on academic data from semester 1, the attributes of student data in semester 2 including student status in semester 2 (Status_2), IP in semester 2 (IPS_2, IPK_2), number of credits in semester 2 (Jum_sks_2) and student registration number (No_Daftar) are removed. This process leaves student data in semester 1 including NIM, study program, student status in semester 1 (Status_1), IP in semester 1 (IPS_1, IPK_1) batch, ICE Placement Test score (IC) and ICE level (level). The following is the source code to create a new dataframe with the attributes that will be used for the analysis.

Program 3. Data Reduction

```
akm_ice_df = akm_ice[["NIM", "Program_Studi", "IPS_1", "IPK_1", "Status_1",
    "Angkatan", "IC", "level"]]
```

Data Transformation

Data transformation is a stage in data preprocessing that changes or converts data into a form of data that is more appropriate to needs. In this study, the Placement ICE (IC) value data has different scale values. The values for the batches before the 2020 batch (2017, 2018, and 2019) are values 1 - 70 while the 2020 and 2021 batches are values 1 - 50. Therefore, a transformation is carried out in the form of data normalization with the aim of changing the values to the same scale without damaging the differences in the range of values in a dataset. (Kappal, 2019). The normalization technique used is min-max scaling by determining the minimum and maximum range of placement test ICE (IC) values based on the batch. The results of this data normalization are stored in a new column named the 'score' column..

Program 4. Data Transformation

```
def min_max_scaling(x, min_val, max_val, new_min, new_max):
return ((x - min_val) / (max_val – min_val)) * (new_max – new_min) + new_min
skor_min_max = {
    2017: (1, 70),
    2018: (1, 70),
    2019: (1, 70),
    2020: (1, 50),
    2021: (1, 50)}
drop_df['skor'] = drop_df.apply(lambda row: min_max_scaling (row['IC'],
    skor_min_max[row['Angkatan']][0], skor_min_max[row['Angkatan']][1], 1, 100),
    axis=1)
print(drop_df)
```

ANOVA Model Creation

The principle of the ANOVA test is to analyze whether there is a significant variation or difference among variables values (Ghozali, 2011). The ANOVA test was conducted by comparing the cumulative IP 1 (IPK_1) with the ICE level (level) to determine whether there was a significant difference between the two attributes. (Berenson , Levine , & Stephan, 2020) (Albright & Winston, 2019). The parameters used are IPK_1 as the dependent variable and level as the independent variable.

Program 5. ANOVA Test

```
anova = ols('IPK_1 ~ level', data = akm_ice).fit()
tabel_anova = sm.stats.anova_lm (anova, typ=2)
print(tabel_anova)
```

Pairwise T-Test Model Creation

If the ANOVA test results show a difference, the test is continued to find out which pairs of groups are significantly different. Further testing is done using the pairwise t-test. (Berenson , Levine , & Stephan, 2020). The results of the pairwise t-test are evaluated based on the p-value compared to the alpha value (α). If α > p-value then the pair of groups are significantly different. In this study, the pair of groups is the ICE level so that the model uses the level parameter.

Program 6. t-test pairwise

```
uji = anova.t_test_pairwise('level')
result = uji.result_frame
print(result.transpose())
```

RESULTS AND DISCUSSION

Student Profile

The data used in the study is communicated visually to make it easier to understand, analyze, and draw conclusions. Figure 1 shows a graph of the number of students per class in the Informatics and Information Systems department. The number of students peaked in 2018.

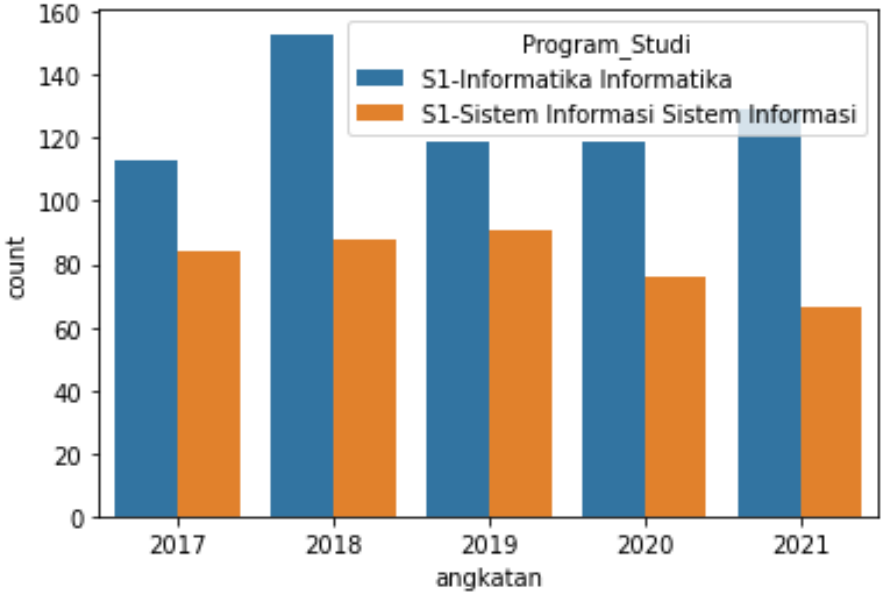

Figure 1. Number of Informatics and Information Systems Students

The descriptive IPS_1 of FTI students in each class can be seen in table 1. The descriptive IPS_1 of the Informatics and Information Systems study program can be seen in tables 2 and 3.

Table 1. Descriptive IPS_1 of FTI Students in Each Batch

Angkatan		2017	2018	2019	2020	2021
IPS_1	count	186.000000	230.000000	203.000000	194.000000	185.000000
	mean	2.833710	2.808261	2.866256	3.253041	2.985297
	std	0.827655	0.737098	0.651166	0.687829	0.956726
	min	0.150000	0.330000	0.300000	0.900000	0.170000
	25%	2.427500	2.435000	2.550000	3.012500	2.570000
	50%	3.045000	3.000000	2.930000	3.465000	3.370000
	75%	3.400000	3.327500	3.325000	3.787500	3.680000
	max	4.000000	3.960000	3.900000	4.000000	4.000000

Table 2. IPS_1 of Informatics Students in Each Batch

Angkatan		2017	2018	2019	2020	2021
IPS_1	count	109.000000	144.000000	114.000000	118.000000	120.000000
	mean	2.682752	2.620833	2.699561	3.169576	2.809750
	std	0.821022	0.700810	0.683083	0.720499	0.918198
	min	0.150000	0.350000	0.300000	0.900000	0.170000
	25%	2.240000	2.270000	2.365000	2.995000	2.435000
	50%	2.670000	2.620000	2.715000	3.405000	3.160000
	75%	3.340000	3.162500	3.185000	3.685000	3.432500
	max	3.990000	3.960000	3.850000	4.000000	3.950000

Table 3. IPS_1 of Information Systems Students in Each Batch

Angkatan		2017	2018	2019	2020	2021
IPS_1	count	77.000000	86.000000	89.000000	76.000000	65.000000
	mean	3.047403	3.122093	3.079775	3.382632	3.309385
	std	0.794114	0.691610	0.540412	0.616167	0.948411
	min	0.170000	0.330000	0.720000	1.280000	0.500000
	25%	3.000000	3.040000	2.880000	3.020000	3.380000
	50%	3.270000	3.225000	3.150000	3.620000	3.730000
	75%	3.450000	3.487500	3.380000	3.850000	3.830000
	max	4.000000	3.950000	3.900000	4.000000	4.000000

Figure 2 shows the ICE level for all students. In general, the basic English ability of new students is very poor. Very few students (4.9%) have standard English ability that is considered ready to take English language courses and do not need to take the ICE program. The majority of students are at levels 1 and 2. The details of the distribution of ICE levels per batch in both study programs are shown in Table 4.

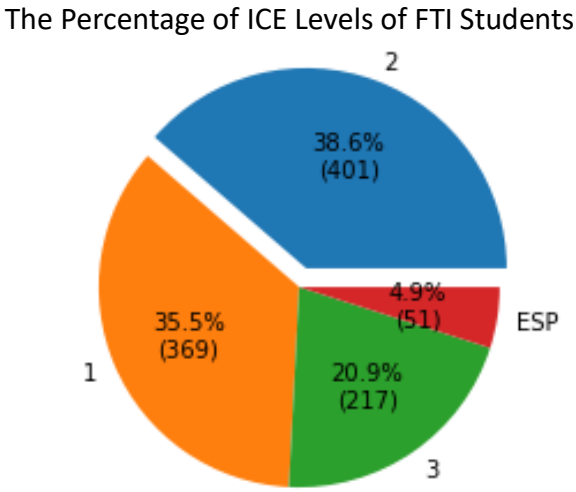

Figure 2. ICE Level for New FTI Students

Table 4. Distribution of Number of Students per ICE Level

Batch/Department		ICE Level 1	2	3	ESP
2017	INF	45	42	15	11
	SI	55	22	5	2
2018	INF	56	52	28	17
	SI	52	23	10	3

2019	INF		46	38	25	10
		SI	54	26	9	2
2020	INF		16	60	42	1
		SI	11	47	19	1
2021	INF		21	58	47	3
		SI	13	33	19	1

The Influence of ICE Level on GPA

Boxplot figure 3 shows the GPA of semester 1 students (IPS_1) of both study programs grouped based on their ICE level..

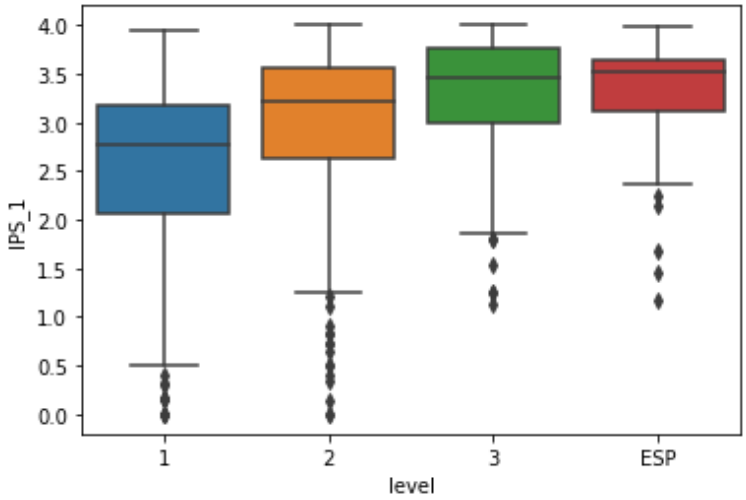

Figure 3. FTI Student IPS_1 at Each ICE Level

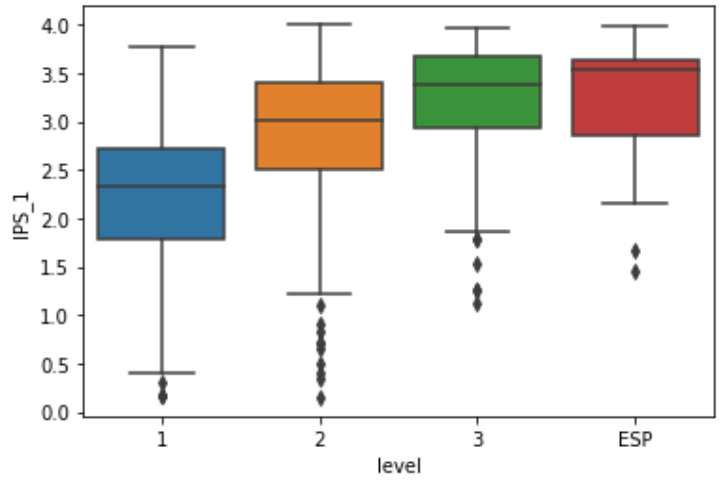

Figure 4. Informatics Student IPS_1 at Each ICE Level

The boxplot of Informatics and Information Systems department students is shown in Figures 4 and 5. It appears that in both study programs, the higher the ICE level, the higher the IPS. Students with an ICE level of 3 or ESP tend to have an IPS > 2 with fewer outlier data.

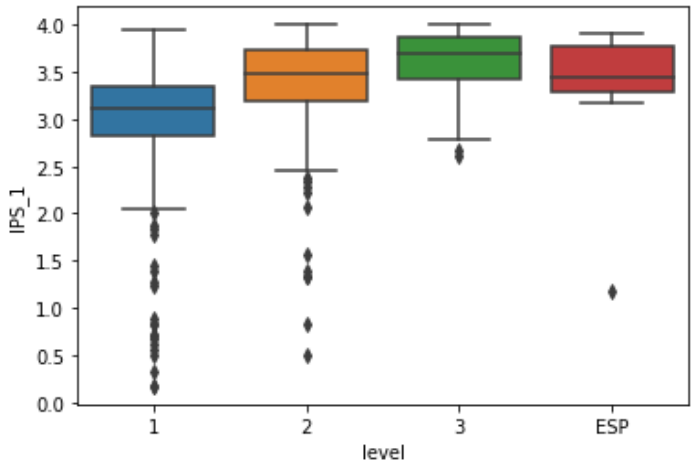

Figure 5. Information System Student IPS_1 at Each ICE Level

ANOVA Testing

One way ANOVA test was conducted to determine whether there was a significant difference in IPS_1 between the 4 ICE groups in all FTI students. The results are shown in Figure 6. The PR(>F) value approaching 0 indicates that there is a significant difference in IPS_1 between the four ICE levels in all FTI students..

	sum_sq	df	F	PR(>F)
level	92.857688	3.0	53.189037	6.704029e-32
Residual	581.935001	1000.0	NaN	NaN

Figure 6. FTI Students' ANOVA Test Results

Details of F and PR (>F) values for students in each study program and each batch are shown in table 5. With α = 5% (0.05) it appears that there is a significant difference in IPS_1 between different ICE groups for all batches and all study programs. The greatest significance occurred in FTI students in the 2020 batch where lectures were conducted online due to the Covid-19 pandemic..

Table 5. PR(>F) Values of ANOVA Test for Both Departments and Batch

Batch	F	PR(>F) %
FTI 2017	4.30269	0.0058
FTI 2018	6.58115	0.0003
FTI 2019	9.15747	0.0000
FTI 2020	23.02283	0.0000
FTI 2021	15.20367	0.0000
Informatics all batch	59.1633	0.0000
Inf.System all batch	21.3283	0.0000

Since there is a significant difference in IPS_1, the next step is to test which pairs of ICE levels are significantly different. The Python output for the test on all FTI students is shown in Figure 6..

	sum_sq	df	F	PR(>F)
level	92.857688	3.0	53.189037	6.704029e-32
Residual	581.935001	1000.0	NaN	NaN

	2-1	3-1	ESP-1	3-2	ESP-2	ESP-3
coef	0.470972	0.775103	0.71518	0.304132	0.244208	-0.059923
std err	0.056126	0.066065	0.117395	0.064798	0.116687	0.12178
t	8.391395	11.732401	6.092082	4.693517	2.092855	-0.492064
P>|t|	0.0	0.0	0.0	0.000003	0.036614	0.622782
Conf. Int. Low	0.360834	0.645461	0.484811	0.176976	0.015229	-0.298897
Conf. Int. Upp.	0.581109	0.904746	0.945549	0.431288	0.473187	0.17905
pvalue-hs	0.0	0.0	0.0	0.000009	0.071887	0.622782
reject-hs	True	True	True	True	False	False

Figure 6. Pairwise t-Test on All FTI Students

The conclusion of the test is shown in the bottom row of Figure 6. Significant differences occur in the True value column, namely between levels 1-2, 1-3, 1-ESP and 2-3. Details of the t-Test results for each student class, both for all FTI students and students in each study program, are shown in Table 6. The numbers in Table 6 show the ICE level with IPS_1 which are significantly different (α = 5%). It appears that in the majority of cells, a significant difference appears at ICE level 1. So in general, students with poor English skills will tend to have a semester 1 GPA (IPS 1) that is much different from students who enter the ICE level above it.

Table 6. Pairwise t-test Results for Each Departement and Batch

Batch	FTI		Informatics		Information System
2017	1-ESP		1-ESP	2-ESP	1-2
2018	1-3 1-ESP		1-2 1-3 1-ESP	2-3 2-ESP	
2019	1-2 1-3 1-ESP		1-2 1-3 1-ESP		
2020	1-2 1-3 1-ESP	2-3	1-2 1-3	2-3	1-2 1-3
2021	1-2 1-3 1-ESP	2-3	1-2 1-3		1-2 1-3
All Batch	1-2 1-3 1-ESP	2-3	1-2 1-3 1-ESP	2-3 2-ESP	1-2 1-3

Discussion

The results of the ANOVA test and the Boxplot graph show that there is a significant difference in semester 1 IPS between students with poor English skills (ICE level 1) and students with good basic English skills (ICE level 3 or ESP). The significant difference in IPS in the Informatics

study program occurs more variedly at various levels compared to the Information Systems study program. However, in both study programs, significant differences in IPS always occur in students with low English skills (level-1) with the level above it. Differences in IPS in students with moderate English skills (level 2) with the level above it occur in fewer cases. This fact shows a positive relationship between basic English skills and the results of students' semester 1 studies. These results are in line with the results of research in Thailand (Rudd & Honkiss, 2020), Malaysia (Suyansah & Gabda, 2020), Indonesia (Desiani, Yahdin, & Rodiah, 2020), as well as in Canada (Neumann, Padden, & Mcdonough, 2019).

CONCLUSION

The results of the ANOVA test show that the basic English language skills of Information Technology department students are closely related to the results of semester 1 GPA. The results of the Pairwise t-test show that the main significance of the majority occurs in students with poor English skills (level 1) with students with high English skills (level 3 or ESP). The pattern of the relationship between basic English language skills and semester 1 study results is not linear, so it is necessary to try predictions using the Neural Network model as a follow-up study.

REFERENCES

Albright, C. S., & Winston, W. L. (2019). *Business Analytics: Data Analysis & Decision Making, 7th Edition.* Cengage Learning.

Berenson , M. L., Levine , D. M., & Stephan, D. F. (2020). *Basic Business Statistics: Concepts and Applications, 14th edition.* England: Pearson Education Limited.

Desiani, A., Yahdin, S., & Rodiah, D. (2020). Prediksi Tingkat Indeks Prestasi Kumulatif Akademik Mahasiswa dengan Menggunakan Teknik Data Mining. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK), 7*, 1237-1244. doi:https://doi.org/10.25126/jtiik.2020722493

Fenton-Smith, B., Humphreys, P., & Walkinshaw, I. (2015). Implementing a University-Wide Credit-Bearing English Language Enhancement Programme: Issues Emerging from Practice. *Studies in Higher Education, 42*(3), 1-17. doi:doi:10.1080/03075079.2015.1052736

Ghozali, I. (2011). In *Aplikasi Analisis Multivariate dengan Program SPSS* (Vol. 5, pp. 61-88). Semarang: Badan Penerbit Universitas Diponegoro.

Han, J., Kamber, M., & Pei, J. (2011). In *Data Mining: Concepts and Techniques* (pp. 93-94). Elsevier Science.

Iswindarti, I., Rufii, & Hartono. (2020). Pengaruh Penerapan Metode Learning by Conversation dan Kepercayaan Diri Terhadap Prestasi Belajar Mata Pelajaran Bahasa Inggris. *Jurnal Pembangunan Pendidikan Fondasi dan Aplikasi, 8*(2), 116-125. doi:10.21831/jppfa.v8i2.38823

Kappal, S. (2019). Data Normalization Using Median & Median Absolute Deviation (MMAD) based Z-Score for Robust Predictions vs. Min-Max Normalization. *London Journal of Research in Science: Natural and Formal, 19*(4), 39-44. doi:doi:10.13140/RG.2.2.32799.82088

Matthew, C. K., & Widjaja, Y. (2022). Hubungan Antara Burnout dengan Academic Procrastination pada Mahasiswa Fakultas Kedokteran Universitas Tarumanagara Tahap Akademik. *Jurnal Pendidikan Dan Konseling (JPDK), 4*(6), 7431–7440. doi:https://doi.org/10.31004/jpdk.v4i6.9515

Muhardi. (2004). Kontribusi Pendidikan dalam Meningkatkan Kualitas Bangsa Indonesia. *Jurnal Sosial dan Pembangunan, 20*, 478-492. doi:https://doi.org/10.29313/mimbar.v20i4.153

Neumann, H., Padden, N., & Mcdonough, K. (2019). Beyond English Language Proficiency Scores: Understanding The Academic Performance of International Undergraduate Students during The First Year of Study. *Higher Education Research and Development, 39*(2), 324-228. doi:doi:10.1080/07294360.2018.1522621

Rudd, M., & Honkiss, L. (2020, February). Analysing the Correlation between English Proficiency and Academic Performance among Thai University Students. *Athens Journal of Education, 7*(1), 122-138. doi:https://doi.org/10.30958/aje.7-1-6

Sherlywati, Junita, I., Kristine, F., & Wisamtamma, E. (2021). Perbedaan Persepsi Bimbingan Akademik Mahasiswa Pada Sebuah Perguruan Tinggi Swasta di Indonesia. *Jurnal Manajemen Maranatha, 20*(2), 119-134. doi:https://doi.org/10.28932/jmm.v20i2.3133

Sihotang, S. F. (2022). Analisis Hasil Belajar Mahasiswa Universitas Potensi Utama Menggunakan Two Way ANOVA. *MES: Journal of Mathematics Education and Science, 8*(1), 173-179. doi:DOI: https://doi.org/10.30743/mes.v8i1.6115

Sunandar, Alvarez, A. I., & Cardozo, C. A. (2024). The Influence of the Use of Role Playing Methods and Interest in Learning on the English Speaking Ability. *Tekno-Pedagogi Jurnal Teknologi Pendidikan, 14*(1), 16-24. doi: https://doi.org/10.22437/teknopedagogi.v14i1.32519

Suyansah, Q., & Gabda, D. (2020). An Analysis of the Effect of English Proficiency towards Students' Academic Performance in University of Malaysia Sabah. *Universal Journal of Educational Research*, 83-88. doi:doi:10.13189/ujer.2020.081510