Gadjah Mada Journal of Psychology (GamaJoP)
Journal Homepage: https://jurnal.ugm.ac.id/gamajop
Email : gamajop.psikologi@ugm.ac.id
ISSN 2407 - 7798 (Online)

Volume **10**
Number 2, 2024
Page: 118–130
DOI:10.22146/gamajop.83491

Received 29 March 2023
Revised 24 July 2024
Accepted 18 September 2024
Published 25 October 2024

Keywords:
academic achievement; differential aptitude test; flanagan aptitude classification test; predictive validity

***Author for correspondence:** Email: r.hidayat@ugm.ac.id

Assessing the Predictive Power of Aptitude Tests on Academic Achievement of Students in Science and Technology Majors

Anindita Dwi Hapsari and Rahmat Hidayat*

Faculty of Psychology, Universitas Gadjah Mada, Indonesia

Abstract

Aptitude tests are used as part of aptitude assessments. The results of these tests can later serve as the basis for providing study major recommendations for clients who wish to pursue their studies at higher education institutions. Two of the aptitude tests used as part of the aptitude assessment are the Differential Aptitude Test (DAT) and the Flanagan Aptitude Classification Test (FACT) which consisted of comprehension, reasoning, patterns, arithmetic, mechanical reasoning, and speed-accuracy subtests. This research intended to know the predictive power of aptitude tests on academic achievement in higher education institutions, especially in the STEM program. Participants in this study were 179 students (111 female, 68 male) at Universitas Gadjah Mada. All participants had taken an aptitude assessment at a psychological service provider and had undergone their first year of study in the STEM programs. Correlation analysis was carried out with the aptitude test as the predictor and GPA as the criterion. This study found that only the numerical ability subtest had a satisfactory correlation coefficient as a predictor of the academic performance in female STEM students in higher education.

As of 2020, there were 4,593 higher education institutions with 29,413 study programs from various scientific fields in Indonesia (Kemenristekdikti, 2019). Study programs in Indonesia are grouped into several scientific fields, including agricultural, social, educational, religious, engineering, economic, health and natural sciences (MIPA), humanities, and arts (Kemenristekdikti, 2019). The alignment between interest and aptitude of individuals in their scientific fields contributes to academic success (York et al., 2015). In determining whether a certain study domain matches the interests and aptitude of each individual, a specific assessment process is needed to obtain comprehensive results. In the field of educational psychology, psychological assessments are widely carried out with the aim of determining the intellectual capacity and identifying the interests and aptitude of students, which can later be used to predict the level of success of students in pursuing their studies (Suwartono, 2020). The results of this psychological assessment can later provide objective information about the student's abilities. Marsidi and Hatta (2019) conducted an interest and aptitude assessment on 208 high school students in Bekasi who were confused about choosing their undergraduate study programs. The results of the psychological examination contain a lot of information about intelligence capacity, personality, attitudes and emotions in undergoing the learning process, interests, and recommendations for study programs at college.

Rostiana and Saraswati (2018) also conducted a psychological assessment to get an overview of the aptitudes and interests of 269 high school students in Yogyakarta who reported difficulty in choosing majors. This assessment was found useful as considerations for students in choosing a major at college. One of the instruments widely used by psychologists and researchers in psychological assessment is the psychological test. The use of psychological has several functions in the assessment process, namely enhancing the psychologist's understanding of the client, providing evidence to support the diagnosis, and complementing evidence when using more than one assessment tool (Weiner, 2013). These functions make psychological tests play a significant role in the assessment process, so they are often used by psychologists or researchers.

An instrument must provide empirical evidence of certain psychometric criteria

to carry out its function in psychological measurement. This empirical evidence is obtained from several studies that will play a role in enforcing evidence-based assessment (EBA). EBA itself refers to the use of empirical studies and theories in guiding the selection of test instruments to be used for specific assessment purposes (Hunsley & Mash, 2007). EBA is part of the Evidence-Based Psychological Practice (EBPP) policy implemented by American Psychological Association (APA) (2006). This policy encourages more effective psychological practice by applying the principles of psychological assessment, case formulation, therapeutic relationships, and psychological interventions that have empirical supports (American Psychological Association (APA), 2006).

Understanding assessment tools begins with knowing their specific purposes, which is closely related to the validity of the instruments and the accuracy of the test results (American Psychological Assosiation, 2020). Information from empirically tested assessment tools will help psychologists carry out the basic goals of psychology, namely understanding the differences between individuals, groups, and systems. This basic goal is divided into four stages: describe, predict, understand, and influence (Periantalo, 2015). One of these goals is to predict, meaning psychologists can predict psychologically related events that may occur in the future for clients (Daulay, 2014). Therefore, it is important for assessment tools to have empirically tested predictive validity to carry out their predictive function properly and in accordance with EBA policies.

Predictive validity, in a broader sense, refers to the predictive power of a measuring instrument against any criteria within a certain time interval (Anastasi, 1976). This is different from concurrent validity, where the predictor variable and the criterion variable are measured at the same time. Predictive validity is especially important in instruments that measure future performance, e.g. employee selection tests and college entrance examination. According to Siswanto (2014), several factors influence the predictive validity of a measuring instrument, namely its reliability and the time interval between the measurement of the predictor and criterion. The time interval between the assessments of the predictor and criterion is crucial because a prediction tends to be less accurate if the interval is too long. A prolonged interval between the predictor and criterion assessments can result in low correlations due to many factors affecting a person's condition during that period (Siswanto, 2014).

Research on predictive validity is still widely carried out in psychology. Predictive validity was examined on a language and behavioral development screening instrument for children under 5, providing satisfactory predictive power to forecast children's future language and behavioral development (Sim et al., 2019). Another study aimed to determine the predictive validity of the instrument for measuring perception of challenges and its influence on career aspirations (Krannich et al., 2022). The predictive validity of the Kessler Psychological Distress Scale (K10) was also studied to find its effect on the mental well-being of children and adolescents (Smout, 2019). In education, there was a study on how achievement tests, general ability tests, and high school grades can predict the academic achievement of engineering students (Abdelfattah et al., 2021). However, this study did not specify which abilities had the greatest influence on the academic achievement.

Aptitude has become a psychological construct that is often studied in the field of psychology. Aptitude or talent can be interpreted as the capacity of an individual's ability to acquire a skill in a particular field through training (American Psychological Assosiation, 2022). The principle of individual differences also applies to talent, meaning that each individual has their own unique talent. At first glance, human talent and intelligence are different things, but they are interrelated mental abilities in humans. The term talent often refers to one of various different characteristics, independent of one another. This approach then aims to measure various separate skills, similar to the Cattell-Horn-Carroll (CHC) theory and other modern intelligence theories. However, aptitude tests are more likely to be designed for use in education and research (Nugba & Quansah, 2021).

Aptitude itself has several definitions. According to Freeman (1976), aptitude is a combination of characteristics that indicate an individual's capacity to acquire (with training) some knowledge, skills, or a series of organized responses, such as the ability to master language, become a musician, or perform mechanical work. Chauhan (2009) described aptitude as an individual's ability for a certain type of activity, as well as the capacity to acquire proficiency under appropriate conditions, namely their current potential as expressed by performance on a particular test with predictive value. Aptitude itself shows an individual's potential or latent capacity that can be developed to obtain certain proficiency. Based on this statement, it can be interpreted that talent can predict achievement in a particular field and can be measured directly with a specific assessment tool (Woodworth, 2010). The measurement tool commonly used to identify talent is often referred to as an aptitude test.

An aptitude test is designed to measure a person's potential ability in a certain type of activity within a limited range (Freeman, 1976). Aptitude tests are often presented as multiple battery tests, where several subtests in different fields are given to the testee. These subtests can measure various abilities, such as numerical ability, spatial ability, numerical reasoning, and speed and accuracy (Setiawati, 2020). Therefore, aptitude tests can provide a cognitive profile of an individual's areas of strength and weakness (Anastasi, 1976).

Aptitude tests often used by practitioners and researchers include the Differential Aptitude Test (DAT) and the Flanagan Aptitude Classification Test (FACT). DAT was developed by Bennet et al. (1947) in the United States (Bennet et al., 1947), while FACT was first published by Flanagan (1953)

Several previous studies have used DAT and FACT to identify aptitudes and interests. Novalina et al. (2021)

used DAT to explore the interests and aptitudes of students at an institution in Medan. DAT was also used to assess whether the students' aptitudes and interests match their chosen majors at at private vocational schools in Surakarta (Hertinjung et al., 2020). Ardiyaning (2012) conducted research on the development of student potential using online aptitude test with the DAT and FACT comprehension subtests as measuring instruments. Nurcahyo et al. (2012) used DAT and the comprehension subtest from FACT on grade 12 students in Mentawai Islands, West Sumatra, in their study.

Aptitude tests can be used for many purposes, including predicting a person's future performance in various contexts, such as academic achievement (Asrijanty, 2014). Therefore, the predictive ability of the assessment tool is very important. The more precise and accurate an aptitude test is in predicting a student's academic achievement in a specific field, the better the quality of the test (Setiawati, 2020). The accuracy and precision of the test in predicting future achievement can be determined by examining the predictive validity of the assessment (Azwar, 2018). Predictive validity can be determined by calculating the correlation between the test results as predictors and the criteria (Azwar, 2018), in this case the criteria would be academic achievement.

The predictive validity of aptitude tests such as the DAT and FACT has been widely studied. Doppelt and Bennett (1951) conducted a longitudinal study by presenting the DAT to high school students in the United States. The study found that the verbal reasoning and language use subtests had the highest correlation with students' academic performance over three years, while the clerical speed and accuracy subtest had the lowest correlation. Additionally, Layton and Swanson (1958) studied the correlation of the DAT in high school students in the United States and found that only the verbal reasoning and numerical ability subtests had significant correlation with academic achievement. Nijenhuis et al. (2000) found that aptitudes measured by DAT showed strong predictive value for achievement in mathematics and Dutch language among immigrant students. However, the study did not specifically mention which subtests had a significant influence on each subject. Research conducted by Setiawati (2020) found that aptitudes measured using the DAT as a whole contributed to the success of psychology students by 41.1%. However, separately, only the verbal and numerical subtests had a significant influence on the academic achievement of psychology students.

Understanding which subtests play a significant role in predicting students' academic achievement is crucial to know what aptitudes are important in a particular field of study. Each field of study has unique core competencies that differentiates it from other fields. The difference in core competencies means the required aptitudes also differ. In science, technology, engineering, and mathematics (STEM), the cognitive skills needed include information processing and management, as well as critical, creative, and computational thinking (UNESCO International Bureau of Education and Soo Boon Ng, 2019). The science and technology cluster in Indonesia covers various fields, including technology, engineering, agriculture, health, and other fields related to the natural sciences (Pradana et al., 2021). Competencies in these fields are certainly different from the competencies needed in social sciences and humanities.

The numerical and verbal reasoning subtests of the DAT have significant predictive power on the academic achievement of grade 11 high school students studying STEM (Santos & Boyon, 2020). In line with this study, the combination of numerical and verbal reasoning aptitudes can also predict academic achievement among college students (Corengia et al., 2013). Haciomeroglu (2016) stated that verbal reasoning ability is the most potential contributor to students' performance in calculus. Rockinson-Szapkiw et al. (2014) said that although both numerical and verbal reasoning ability significantly predicted students' academic performance, numerical ability had greater power in predicting graduate student performance than verbal ability.

In Indonesian universities, engineering is the second most studied field and the most studied one in nine provinces, namely Jakarta Metropolitan Area, Banten, West Java, Central Java, Special Region of Yogyakarta, Bangka Belitung, East Kalimantan, Riau Islands, and South Sumatra (Kemenristekdikti, 2019). In addition, STEM majors are also often highly selective (Universitas Gadjah Mada [UGM], 2022). The selectivity of study programs at UGM in for incoming students in 2020 can be seen in Table 1. 1 also shows that in 2022, the medicine is the study program with the lowest acceptance rate at UGM, at 1.18%. The information and technology major is also a highly selective study program, having an acceptance rate below 2.5% from 2020 to 2022. The data suggests that many students want to advance their studies in the STEM field. However, until now there has been no research on the predictive validity of aptitude test for student academic achievement in science-related study programs in Indonesia.

Based on the background that has been described, this study aimed to explain the ability of the FACT and DAT tests in predicting academic achievement of students majoring in STEM. The hypothesis proposed in this study is that the FACT and DAT comprehension subtests consisting of subtests of reasoning, patterns, arithmetic, mechanical reasoning, and clerical speed and accuracy can predict the academic achievement of students in STEM programs.

Methods

The predictor variables in this study were DAT and FACT scores. The criterion used in this study was academic achievement measured using the Grade Point Average (GPA). Academic achievement is defined as the final result of the learning process that students have undergone over a certain period and is represented in the form of specific numbers or symbols (Suryabrata, 2006). GPA,

Table 1
Acceptance Rates of Several Study Programs at Universitas Gadjah Mada

Major	Acceptance Rate (%)		
	2022	2021	2020
Korean Language and Culture	3.57	2.00	1.54
Health Nutrition	2.44	2.70	1.20
Actuarial Science	3.03	2.78	1.09
International Relations	1.92	1.82	0.87
Computer Science	1.89	2.17	1.03
Communication Sciences	1.30	1.45	0.68
Medicine	1.18	1.49	0.88
Management	1.67	1.56	0.88
Management and Public Policy	2.38	2.86	1.00
Tourism	2.94	2.38	1.25
Psychology	2.44	2.38	1.43
Statistics	2.33	2.63	1.16
Information Technology	2.13	1.92	1.19

as an indicator of academic achievement, refers to the level of success in a particular field of study as evidenced by competence at a certain level.

The instruments used in this study were the DAT test, the comprehension subtest of FACT, and GPA. The DAT test was developed by (Bennet et al., 1947), based on the primary mental abilities (PMA) model by L. L. Thurstone. The DAT test used in this study consisted of five subtests: reasoning (A3), arithmetic (A5), patterns (B3), mechanical reasoning (C4), and clerical speed and accuracy (D4). All DAT subtests are "power tests" except for the clerical speed and accuracy test, which is a "speed test" (Anastasi, 1976). The reliability of the DAT subtests that have been adapted to the Indonesian language varies from 0.73 to 0.92. The reliability of the reasoning, arithmetic, pattern, and mechanical reasoning subtests are 0.858, 0.850, 0.862, and 0.737, respectively (Setiawati et al., 2018). Meanwhile, the reliability of the clerical speed and accuracy subtest is 0.77-0.93 for men and 0.84-0.91 for women (Bennet et al., 1952).

The next instrument is the comprehension subtest from FACT. The FACT comprehension subtest is the eighth subtest in the FACT series. In the original version, this test is called the judgement and comprehension subtest. This subtest measures an individual's ability to read and understand, then assess the logic of reasoning in decision by understanding a practical situation. The reliability of the comprehension subtest that has been adapted into Indonesian is 0.66 (Suramto et al., 1996).

The distribution of items for DAT and FACT can be seen in Table 2. Meanwhile, the results of the assessment of graduate learning outcomes at the end of the study program are stated in the GPA (Kementerian Pendidikan dan Kebudayaan Republik Indonesia, 2014). The GPA is quantified by adding the multiplication of the grades of each course taken and the credits (SKS) of the relevant course, divided by the number of credits of the courses that have been taken.

This research received ethical clearance from the Research Ethics Committee of the Faculty of Psychology at UGM (clearance number 1669/UN1/FPSi.1.3/SD/PT.01.04/2022).

The data in this study are secondary in the form of aptitude tests (one FACT subtest and five DAT subtests) done by incoming students of STEM majors at Universitas Gadjah Mada year 2020. Criteria data were in the form of the grade points of semester 1, grade point of semester 2, and grade point average for the first year of study. Data were obtained from the psychology bureau and academic divisions of STEM at UGM, namely the Faculty of Biology; Faculty of Pharmacy; Faculty of Geography; Faculty of Medicine, Public Health and Nursing; Faculty of Veterinary Medicine; Faculty of Dentistry; Faculty of Forestry; Faculty of Mathematics and Natural Sciences; Faculty of Agriculture; Faculty of Animal Science; Faculty of Agricultural Technology; Faculty of Engineering; and the Vocational School. The data collection procedure is explained as follows:

Before data collection was carried out, the researcher applied for permission to conduct research from the psychology bureau and the academic divisions of the target faculties. Then, the psychology bureau together with the researcher asked for permission to conduct research on users of the psychology bureau services whose data would be included in this study. After obtaining permission, the researcher then collected data from participants who took the aptitude test at the psychology bureau from September to October 2019. The participants chosen were those at that time because the data had complete predictor scores with a sufficient number to conduct predictive validity research. After obtaining the names of the participants, the researcher screened participants who were accepted into the STEM majors at UGM. The researchers then collected data on the first-semester grade point average (GPA), second-semester GPA, and first-year cumulative GPA from the academic divisions of target faculties.

Correlation analysis was used in this research with the aim of knowing the relationship between predictors and criteria. In this study, the predictors were the FACT comprehension and the DAT subtests (reasoning, arithmetic, pattern, mechanical reasoning, and clerical speed and accuracy). Meanwhile, the criterion variables in this study were first-semester GPA, second-semester GPA, and first-year CGPA. Data analysis was conducted using Jamovi software version 2.3.16.0.

Results

The participants included in this study were class of 2020 at UGM who had taken the first year of college and had taken the FACT and DAT tests. The total FACT and DAT test participants that year were 842 students consisting of 735 students majoring in natural sciences (IPA), 101 students majoring in social sciences (IPS), and six students majoring in language. The distribution of test participants can be seen in Table 3.

Based on Table 3, 321 participants were male (280

Table 2
Item Distribution of FACT and DAT Subtests

Subtest	Code	Total Items	Working Time	Score Range
Comprehension	A1	24	30 minutes	0 – 22
Reasoning	A3	50	30 minutes	0 – 50
Arithmetic	A5	40	30 minutes	0 – 36+
Pattern	B3	40	30 minutes	0 – 100
Mechanical Reasoning	C4	68	30 minutes	0 – 68
Clerical Speed and Accuracy	D4	100	3 minutes	0 – 100

Table 3
Prospective Students Based on Major in High School

High School Major	Male	Female
Natural Science	280	455
Social Science	40	61
Language	1	5

majoring in natural science, 40 in social science, and 1 in language) and 521 were female (455 majoring in natural science, 61 in social science, and 5 in language). After a screening process, it was found that there were 179 students (111 female students and 68 male students) who were accepted into UGM's STEM programs. The students who were accepted into the STEM programs all came from natural science majors. The distribution of participants in the study can be seen in Table 4.

Table 4
Distribution of Research Participants (N=179)

Faculty	Male	Female
Faculty of Biology	1	2
Faculty of Pharmacy	2	7
Faculty of Geography	2	5
Faculty of Medicine, Public Health and Nursing	6	30
Faculty of Veterinary Medicine	1	3
Faculty of Dentistry	0	14
Faculty of Forestry	2	1
Faculty of Mathematics and Natural Science	10	4
Faculty of Agriculture	4	7
Faculty of Animal Science	5	2
Faculty of Agricultural Technology	4	9
Faculty of Technology	31	26
Vocational School	0	1

Based on the data in Table 4, the distribution of research participants is not balanced. The proportion of male participants is around 38% of the total sample. This proportion gap is also seen in some faculties, such as the Faculty of Dentistry where there are no male participants. This proportion gap can lead to an issue in computational statistics due to the possibility of sampling error.

Data Descriptive Results of Predictor Variables

The predictor variables in this study were students' FACT comprehension and DAT scores. DAT test consists of a reasoning subtest, arithmetic subtest, pattern subtest, mechanical reasoning subtest, and clerical speed and accuracy subtest. Score 1 was given for the correct answer and 0 for the wrong answer. The FACT comprehension and DAT subtests were calculated according to the guidance, which might differ from one subtest to another. After obtaining the raw score, the raw score was converted into a weighted score using norms from the Faculty of Psychology of UGM. Seen in Table 5

As shown by Table 5, the number of participants decreased from the previous 842 prospective students to 179 accepted students. In the prospective student group, the highest average score was A1 or comprehension (81.35). Meanwhile, the lowest average (60.29) was for the C4 or the mechanical understanding subtest. The largest variance (655.915) was D4 or clerical speed and accuracy, while the subtest with the smallest variance (306.159) was A1, or the comprehension subtest. The largest standard deviation (25.61) was found in the clerical speed and accuracy subtest and the smallest standard deviation (17.49) was seen in the comprehension subtest.

In the group of accepted students, the highest average score was found in the A3 or reasoning subtest, while the lowest average score was in the B3 or pattern subtest. The largest variance (560.290) was found in the C4 or mechanical reasoning subtest, whereas the smallest variance (165.086) was seen in the A3 or reasoning subtest. The largest standard deviation (23.67) was also found in the C4 mechanical reasoning subtest, and the smallest standard deviation (12.85) was seen in the A3 or reasoning subtest. Seen in table 6

Table 6 shows the differences in the number of participants based on gender. Among male participants, the highest average score (87.80) was seen in the A3 or reasoning subtest. Meanwhile, the lowest average (46.91) was seen in the C4 or mechanical reasoning subtest. The largest variance (593.326) was found in the D4 or clerical speed and accuracy subtest, while the smallest variance (189.858) was seen in the A3 or reasoning subtest. The largest standard deviation (24.35) was found also in the D4 or clerical speed and accuracy subtest, and the smallest standard deviation (13.77) was seen in the A3 or reasoning subtest. This means that the aptitude for speed and accuracy in male participants was more varied, while

Table 5
Descriptive Statistics of the FACT Comprehension and DAT Subtests

Subtest Code	Prospective Students (*N*=842)				Enrolled Students (*N*=179)			
	Mean	Variance	*SD*	Skew.	Mean	Variance	*SD*	Skew.
A1	81.35	306.159	17.49	-1.434	84.56	195.810	13.99	-1.466
A3	81.02	371.863	19.28	-1.312	88.12	165.086	12.85	-1.945
A5	74.03	492.506	22.19	-0.867	84.48	218.082	14.77	-1.474
B3	54.74	515.403	22.70	-0.153	63.14	447.941	21.16	-0.391
C4	60.29	592.983	24.35	-0.387	66.04	560.290	23.67	-0.741
D4	68.73	655.915	25.61	-0.599	73.51	536.072	23.15	-0.824

Table 6
Descriptive Analysis of Predictor Scores in Enrolled Male and Female Students

Subtest Code	Male (*N*=68)				Female (*N*=111)			
hline	Mean	Variance	*SD*	Skew.	Mean	Variance	*SD*	Skew
A1	81.26	270.824	16,45	-1,004	86,57	141,083	11,87	-1,810
A3	87.80	189.858	13.77	-1,648	88.31	151.400	12.30	-2.160
A5	79.55	227.832	15.09	-1.256	87.49	189.980	13.78	-1.869
B3	57.52	437.387	20.91	-0.039	66.57	427.065	20.66	-0.644
C4	46.91	475.395	21.80	-0.028	77.76	252.163	15.88	-1.396
D4	73.04	593.326	24.35	-0.954	73.79	505.857	22.49	-0.734

the aptitude for reasoning was the opposite.

Among female participants, the highest average score (88.31) was also in the A3 or reasoning subtest. Meanwhile, the lowest average (66.57) was seen in the B3 or pattern subtest. The largest variance (505.857) was seen in the D4 or clerical speed and accuracy subtest, while the smallest variance (141.083) was seen in the A1 or comprehension subtest. The largest standard deviation (22.49) was also in the clerical speed and accuracy subtest, and the smallest standard deviation (11.87) was in the comprehension subtest. Like their male counterparts, female students exhibited more variance in clerical speed and accuracy. Meanwhile, the comprehension aptitude showed the least variability than other subtests among female participants.

The criterion variable in this study is academic achievement, represented by GPA. The GPA system at UGM ranges from 0 to 4. Table 7 describes the participants' GPA data.

Table 7
Statistical Description of GPA (N = 179)

Criteria	Min	Max	Mean	Variance	*SD*
GPA[a]1	0.75	4.00	3.42	0.217	0.47
GPA[a]2	0.96	4.00	3.41	0.290	0.54
CGPA[b]	1.04	4.00	3.42	0.229	0.48

[a]Grade Point Average (per semester)
[b]Cumulative Grade Point Average (per study year)

Table 7 first semester GPA and CGPA of the first year had the highest mean of 3.42 each. The first semester GPA also had the largest range of 3.25 (*SD*=0.47). The smallest range was shown in the first-year GPA at 2.96 (*SD*=0.48). The highest standard deviation (0.54) was seen in the second semester GPA, while the lowest was observed (0.47) in the first semester GPA. Therefore, the second semester GPA data was the most heterogeneous, while the first semester GPA was the most homogenous. Seen in table 8

Table 8 shows the descriptive data of participants' GPA based on gender. The average GPA of the accepted students was relatively satisfactory. The largest mean difference (0.159) between male and female participants was seen in the second semester GPA. Meanwhile, the smallest difference (0.076) was seen in the first-year CGPA.

Correlation Analysis Results

The predictive validity of the aptitude test can be seen from the linear correlation coefficient between predictor variables (comprehension subtest, reasoning subtest, arithmetic subtest, pattern subtest, mechanical reasoning subtest, and clerical speed and accuracy subtest) and criterion variables (in this case the first semester GPA, second semester GPA, and CGPA). The results of the simple linear correlation analysis can be seen in Table 9.

The correlation analysis results showed that the highest correlation was between the arithmetic subtest and the second semester GPA (R=0.297; p<0.001). In addition, the lowest correlation is between the reasoning subtest and the second semester GPA (R=-0.002; p>0.001). In Table 6, it can be seen that the aptitude test and GPA had a positive correlation regardless of the significance level, except for the reasoning subtest and the clerical speed and accuracy subtest. In addition, it can also be seen significant correlation with GPA was only observed in the arithmetic subtest. Meanwhile, the reasoning subtest has the lowest correlation coefficient when compared to the other subtests.

Table 8
Descriptive Analysis of GPA in Accepted Students by Gender

GPA	Male (*N*=68)				Female (*N*=111)			
	Mean	Variance	*SD*	Skew.	Mean	Variance	*SD*	Skew.
GPA 1	3.352	0.299	0.547	-2.843	3.454	0.165	0.406	-1.525
GPA 2	3.313	0.299	0.547	-2.093	3.473	0.278	0.527	-2.591
CGPA	3.335	0.281	0.530	-2.492	3.466	0.281	0.439	-2.109

Table 9
Correlation Analysis of FACT Comprehension and DAT Subtests with GPA (N=179)

Criteria	Predictor					
	A1	A3	A5	B3	C4	D4
GPA 1	0.149	-0.026	0.259a	0.109	0.086	-0.108
GPA 2	0.106	-0.002	0.297a	0.093	0.127	-0.149
CGPA	0.133	-0.014	0.293a	0.106	0.114	-0.138

[a] $p<0.001$

The next analysis used in this study was simple linear regression analysis to determine the effectiveness of the arithmetic subtest in predicting academic achievement in students studying STEM. The results of the simple linear regression analysis can be seen in Table 10.

Table 10
Simple Regression Analysis of Arithmetic Subtest and GPA in STEM Students N = 179)

Criteria	R	R^2	β	*t*	*p*
GPA 1	0.259	0.067	0.008	3.565	<0.001
GPA 2	0.297	0.088	0.011	4.131	<0.001
CGPA	0.293	0.086	0.009	4.080	<0.001

In Table 10, it can be seen that the arithmetic subtest had a significant influence on the first-semester GPA, second-semester GPA, and first-year CGPA for students majoring in STEM. The arithmetic subtest had a predictive value that ranged from 6.7 to 8.8% for GPA. Seen in table 11

Table 11 correlation between aptitude test scores and GPA based on gender. This study found that the resulting correlation coefficient was so small that it was close to zero. In addition, there are also some correlation coefficients that have a negative sign. In the male participant group, the correlation coefficients had negative values in the subtests of reasoning, mechanical reasoning, and clerical speed and accuracy. Among male participants, there was no significant correlation between GPA and the arithmetic subtest, although the correlation coefficients were the highest among the subtests. In the female participant group, only the arithmetic subtest had a significant correlation with GPA.

The researchers then conducted a simple linear regression analysis to determine the effectiveness of the arithmetic subtest in predicting academic achievement in female students majoring in STEM. The results of the simple linear regression analysis can be seen in Table 12.

Based on Table 12, it can be seen that the arithmetic subtest had a significant influence on the first-semester GPA, second-semester GPA, and first-year CGPA in female students studying STEM at the university level. The arithmetic subtest accounted for 10-13% of female students' GPA, while the rest was determined by other variables.

Discussion

This study aimed to determine the predictive power of aptitude tests on the academic achievement of students majoring in STEM. The aptitude tests examined in this study have been used by practitioners and researchers to provide recommendations for students who want to continue to pursue higher degree education, particularly in science and technology programs. The STEM programs are known to attract many applicants. Therefore, understanding the predictive power of aptitude tests is crucial to identifying the abilities that determine students' success in these programs.

Predictive power is tested by correlating predictors and criteria to obtain predictive validity coefficients. The correlation coefficient obtained from the analysis shows the relationship between predictors (i.e., aptitude tests) and criteria (i.e., student achievement index). A relatively small correlation coefficient indicates a small size group and low diversity (Azwar, 1996; Furr, 2022; Kurpius & Stafford, 2006). If the group data has a similar variance, albeit small, the scatter plot will show dense data points in one part of the line. It impacts the low correlation coefficient (Allen & Yen, 1979).

The present study found that only the arithmetic subtest had a correlation of more than 0.30 with GPA in the female group, with predictive power of 10-13%. Other subtests did not have satisfactory predictive power as they did not meet the established criterion of 0.3, for both male and female participants. The guidelines by the US Department of Labor, Employment Training, and Administration for interpreting correlation coefficients in predictive validity studies, cited by Emery (Azwar, 2016), state that a validity coefficient smaller than 0.30 is less satisfactory and means that the test in question cannot be valid.

The arithmetic subtest can validly and significantly predict performance in women but not in men. This result contrasts Nurkhafifah et al. (2020) finding that gender did not significantly affect numerical ability. Additionally, a literature review by Spelke (2005) stated that male students showed greater variability in the quantita-

Table 11
Correlation Analysis FACT Comprehension and DAT Subtests with GPA by Gender

Criteria	Gender	*N*	Predictor					
			A1	A3	A5	B3	C4	D4
GPA 1	Male	68	0.147	-0.065	0.155	0.042	-0.077	-0.113
	Female	111	0.115	0.005	0.320a	0.130	0.140	-0.110
GPA 2	Male	68	0.107	-0.061	0.144	0.131	0.010	-0.034
	Female	111	0.062	0.034	0.359a	0.024	0.078	-0.136
CGPA	Male	68	0.132	-0.064	0.154	0.090	-0.034	-0.151
	Female	111	0.090	0.022	0.364a	0.074	0.112	-0.133

[a] $p<0.001$

Table 12
Simple Regression Analysis of Arithmetic Subtest in Female STEM Students (N = 111)

Criteria	*R*	R^2	β	*t*	*p*
GPA 1	0.320	0.103	0.009	3.530	<0.001
GPA 2	0.359	0.129	0.014	4.021	<0.001
CGPA	0.364	0.133	0.012	4.087	<0.001

tive subtest of the SAT exam. However, men and women are comparable when it comes to advanced mathematics in college. This finding is also consistent with this study's data, which shows that the variance in the arithmetic subtest is greater among male participants than female participants, even though arithmetic ability had a greater influence on female students' GPA.

IEA (2007), Indonesian women had higher average algebra scores than men. A study on higher order thinking skills (HOTS) in mathematics found that women performed better in evaluation and analysis, while men were better at creativity (Anggraini et al., 2019). Women also surpass men in certain arithmetic skills, such as complex multiplication, simple subtraction, number comparison, and number sequences (Wei et al., 2012). The arithmetic subtest on the DAT test measures numerical ability, or the ability to understand and use numerical relationships to perform various computational operations quickly and correctly. Numerical evaluation and numerical analysis are needed to excel in this subtest. Therefore, this subtest has a greater influence on women than men.

The arithmetic subtest measures an individual's understanding of numerical concepts. This finding aligns with Layton and Swanson (1958), who found that the numerical ability subtest has a significant effect on academic achievement than other subtests. The numerical ability subtest can also predict performance in fields of study where quantitative thinking is needed, e.g., mathematics, physics, chemistry, engineering, and other areas (Bennet et al., 1952). A study by Dudung et al. (2019) found that numerical ability influences 57.8% of the academic achievement of students in the mechanical engineering department who took the computerized numerical control (CNC) course. This course contains many elements of numerical reasoning, such as mathematics and calculus.

In the first year of college, students in the STEM programs take many basic science courses, such as basic physics, basic mathematics, and basic chemistry. These courses require the understanding of arithmetic concepts, making the arithmetic subtest a good predictor of students' academic achievement in these study programs. Numerical aptitude is one of the strongest predictors of student performance in subjects that require arithmetic skills, such as mathematics (Nozaleda, 2019). Numerical aptitude also has a higher predictive capacity for academic performance than verbal aptitude, spatial orientation, auditory memory, visuomotor skills, and reading and writing maturity (Navarro-Soria et al., 2021).

A longitudinal study by (Vera & Cortes, 2021) stated that the strongest predictor of academic performance was past academic performance, followed by numerical aptitude, verbal aptitude, abstract reasoning, and, emotional self-regulation. The findings in this study align with previous findings, showing that the arithmetic subtest, which measures numerical ability, had the highest correlation coefficient with the academic performance of STEM students compared to other aptitude subtests. Setiawati (2020) also found that the clerical speed and accuracy subtest had low predictive power on the student achievement.

The reasoning subtest measures the ability to recognize patterns in abstract images, while the clerical speed and accuracy subtest measures the speed of an individual's response to simple perceptual tasks (Bennett et al., 1948). In the first year of the science and technology study program, the courses taught are mostly theoretical instead of practical. Meanwhile, the abilities measured by the reasoning subtest and the clerical speed and accuracy subtest are more relevant in activities that require practical skills, which are often needed in workshop classes.

The criteria used in this study were limited to the first year of college, so the resulting predictive power was not very strong. It is possible that these subtests can provide good predictive power in workshop classes, which are mostly taken in the second and third years. An interesting finding from this study is that the correlation between the clerical speed and accuracy subtest and GPA in both male and female participants had negative coefficients. In predictive validity research, negative correlation coefficients cannot determine whether a predictor is valid or not (Azwar & Ancok, 2008). In other words, the

clerical speed and accuracy subtest cannot be used as a predictor of academic performance for STEM students.

The only "speed test" in this study was the clerical speed and accuracy subtest, where the time limit for completion greatly influences the final score of the testee. The faster the participant could complete a task, the better their work performance would be. Tests that use speed to determine performance have been found to have poor predictive validity (Dodonova & Dodonov, 2013). The time limit on tests that measure cognitive performance is less predictive because there are many irrelevant factors in play, such as the administration method (paper-and-pencil or computer-based) and testee behavior (item skipping or random responses) (Lu & Sireci, 2007). Gernsbacher et al. (2020) stated that aptitude tests with time limits have several shortcomings: they are less valid, less reliable, less inclusive, and less fair because of individual differences. Therefore, this type of test cannot be used to measure aptitude.

Another interesting finding in the study is that female participants recorded a positive correlation between the reasoning subtest and GPA, while male participants did not. The reasoning subtest measures abstract reasoning, which assesses the ability to understand relationships in abstract images and principles of nonverbal designs (Bennet et al., 1952). Abstract reasoning requires significant critical thinking skills. A study conducted by Mawaddah et al. (2018) found that women excel in more critical thinking skill indicators, namely analytical and evaluation (Anggraini et al., 2019), which also make women better at understanding relationships in abstract image patterns, as measured by the DAT reasoning subtest.

In this study, although the correlation between the reasoning subtest and academic achievement in female students was higher than in male students, the reasoning subtest coefficient was not yet significant enough to predict academic achievement. Other subtests besides the arithmetic subtest also cannot be good predictors of academic achievement due to their low correlation coefficients. This might be because many factors influence students' academic achievement. Academic achievement is not only influenced by cognitive factors measured by aptitude tests, but also by other factors, such as academic motivation, academic support, parent-child relationships, demographics, family structure, educational values, academic expectations from parents, and personal variables (e.g., emotional and social conditions) (Erawati, 2015). Aptitude included in cognitive abilities is a strong factor in predicting academic achievement, but its contribution is less than 20% (Downey et al., 2013). Good academic performance is also influenced by self-management, emotional management, high awareness, and low extraversion (Downey et al., 2013). Curiosity in science, high self-awareness, socioeconomic status, learning style, and self-confidence also affect academic achievement at the same level as the cognitive domain (Von Stumm et al., 2011).

This low predictive power can also be attributed to the age of the test instrument. The aptitude tests used as predictors in this study were the FACT and DAT tests, which were adapted around the 1980s from the versions developed by Flanagan and Bennett et al. around the 1950s. Thus, the gap between the aptitude test adaptation with this study was about 40 years old. In this 40-year span, there has been no significant revision to the aptitude test used.

According to the Standards for Educational and Psychological Testing, revisions or amendments are needed when new research data or new testing conditions for use and interpretation indicate that the test is no longer optimal. Revisions or amendments are also needed when the test may not fit some of its intended uses, such as when the content or language used in the test has become outdated. Such conditions can affect the validity of interpreting test scores (AERA and APA and NCME, 2014).

In the FACT and DAT subtests, many words used are no longer commonly understood by test participants. This can impact their understanding of the test instructions and affect the validity of the test results. This finding can provide insights into the psychological instruments in Indonesia, especially aptitude assessments, indicating the urgency to evaluate and revise existing instruments.

This study also found that the predictive validity coefficient of aptitude tests tends to decrease each semester in all subtests. This can be attributed to the fact that the time interval for collecting predictor and criterion data greatly affects predictive validity (Siswanto, 2014). The first semester had a closer time gap with the FACT and DAT tests, so the tests had a higher predictive power for the first semester GPA compared to the GPA of the following semesters, which had a greater time gap from the predictor (Fenster et al., 2001; Sternberg & Williams, 1997). Additionally, the courses in the first semester tend to be less difficult. The easier a subject is to study, the more conventional memory and analysis are used, as measured by aptitude tests (Sternberg & Williams, 1997).

The low correlation coefficient seen in this study might also be due to the selection of instruments to measure the criterion variables or academic achievement. In this study, the instruments used as criteria are the grade point semester (GPS) or GPA per semester and the CGPA for the first year. The achievement index is obtained by calculating the scores of the courses that had been taken by students and the credit weight of the courses taken in one semester. This makes it possible for different students to have the same GPA but with different compositions and numbers of subjects. Differences in the methods and assessment standards the lecturers applied and the difficulty level of each course can also lower the predictive validity coefficient (Ramist et al., 1994). The value of courses in higher education is influenced not only by the students but also by the lecturers teaching different courses, even within the same faculty and university (Tierney, 2017).

The initial analysis conducted in this study showed that the correlation between the aptitude test and the GPA of STEM students was low and unsatisfactory. However, after correcting for the effect of distribution restric-

tions on the criterion variables, there was an increase in the correlation coefficient. This increase indicates that the criterion (GPA) data was homogeneous, which played a role in the low correlation coefficient. The GPA of science and technology students was likely homogeneous because the data was obtained from students at UGM, following UGM's rules for evaluating academic performance. Therefore, it did not represent heterogeneous data.

After correcting for the effect of distribution restrictions, only the arithmetic subtest had a satisfactory correlation coefficient. This might be because the language used in the aptitude test is not often used by participants, and the theoretical courses in the first year of science and technology study programs require quantitative skills but not practical skills, unlike the second or third year of university which involves more practical classes.

The limitation of this study is that this study only used the students' GPA in the first year, where they mostly took theoretical courses. Thus, this study could not determine the predictive power of aptitude tests for courses that require a lot of practical skills. In addition, this study was only tested on participants who were students of STEM programs at UGM, meaning that the academic assessment was done following the university's specific policy, so it could not produce more diverse criteria data. This study only focused on predictive validity so that it could only determine whether aptitude tests could be predictors of learning success in STEM majors. This research did not examine the effect size of aptitude test performance in learning success in STEM programs.

Conclusion

This study examined the predictive power of aptitude tests on the academic achievement of STEM students. Research on predictive power examines the correlation of a predictor to a criterion, where data is taken with a certain distance of time after the predictor data is taken. This study used the aptitude test as a predictor and the GPA of first-year STEM students so that there was a one to two years gap between the intake of predictor and criterion data. The predictive power is shown by the correlation coefficient between the aptitude test as a predictor and the GPA of students in the STEM programs as a criterion.

The results in this study showed that only the arithmetic subtest had a satisfactory correlation coefficient to be used as a predictor of academic achievement of female students in science study programs. This could be caused by the relevance of the language used in the aptitude test to the research participants, the difference in the realm of cognitive aptitude between men and women, the content of science courses that require a lot of numerical ability, and not many courses in the field of science that require a lot of practical skills in the first year.

Recommendation

Aptitude tests, especially the FACT and DAT comprehension subtests, can be used as part of the assessment in the field of educational psychology. However, this aptitude test must be accompanied by other assessment tools, such as report cards and intelligence test results, to support professional decision making in recommending study programs. Psychologists can also look at the arithmetic subtest of the DAT test to recommend courses related to the sciences and technology. If the client is already aware of their desire to pursue a STEM-related course, the psychologist can give the DAT arithmetic subtest to the client. This can reduce the resources that must be expended during the assessment process so that the potential for client fatigue can be minimized.

Future researchers are expected to further examine the predictive power of aptitude tests on other study programs to see differences in predictive power between subtests on other study programs. In addition, future researchers can add criteria data, e.g., GPA in the subsequent study years or students from other universities, to see the predictive power of aptitude tests for the entire learning time in college programs.

Declaration

Acknowledgments

The authors would like to thank the Directorate of Research at Universitas Gadjah Mada, the educational institution, and the psychology bureau for contributing to this research.

Author Contributions

ADH and RH designed the research concept. ADH conducted the data collection process, analyzed the data, and wrote the research report. RH developed the research concept, assisted with research permits, supervised the entire research process and wrote the research report.

Conflict of Interest

The authors declare that there is no conflict of interest in the entire process of research, writing, and/or publication of this manuscript.

Funding

The entire research process came from the authors' personal funding.

Orcid ID

Anindita Dwi Hapsari https://orcid.org/0000-0002-4322-6726

Rahmat Hidayat https://orcid.org/0000-0002-1323-2914

References

Abdelfattah, F. A., Obeidat, O. S., Salahat, Y. A., BinBakr, M. B., & Al Sultan, A. A. (2021). The predictive validity of entrance scores and short-term performance for long-term success in engineering education. *Journal of Applied Research in Higher Education*, *14*(4), 1272–1285. https://doi.org/10.1108/jarhe-04-2021-0126

AERA and APA and NCME. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Waveland Press, Inc.

American Psychological Association (APA). (2006). Evidence-based practice in psychology. *American Psychologist*, *61*(4), 271–285. https://doi.org/10.1037/0003-066x.61.4.271

American Psychological Assosiation. (2020). APA Guidelines for Psychological Assessment and Evaluation. https://www.apa.org/about/policy/guidelines-psychological-assessment-evaluation.pdf

American Psychological Assosiation. (2022). APA Dictionary of psychology. https://dictionary.apa.org/aptitude

Anastasi, A. (1976). *Psychological testing (Fourth Edition)* (Fourth, Vol. 9). Macmillan Publishing.

Anggraini, N. P., Budiyono, & Pratiwi, H. (2019). Cognitive differences between male and female students in higher order thinking skills. *Journal of Physics: Conference Series*, *1188*, 012006. https://doi.org/10.1088/1742-6596/1188/1/012006

Ardiyaning, L. (2012). *Seleksi bakat untuk pengembangan potensi siswa berdasarkan Differential Apptitude Test (DAT) berbasis web [Talent selection for student potential development based on web-based Differential Apptitude Test (DAT)]* (Thesis). Universitas Sebelas Maret.

Asrijanty. (2014). Validitas prediktif bakat skolastik dan prestasi belajar sebagai kriteria seleksi masuk Perguruan Tinggi [Predictive validity of scholastic aptitude and learning achievement as selection criteria for university admission]. *Jurnal Pendidikan dan Kebudayaan*, *20*(4), 515–534. https://doi.org/10.24832/jpnk.v20i4.163

Azwar, S. (1996). *Pengantar psikologi inteligensi [Introduction to the psychology of intelligence]*. Pustaka Pelajar.

Azwar, S. (2016). *Dasar-dasar psikometrika II [Basics of psychometrics II]*. Pustaka Pelajar.

Azwar, S. (2018). *Reliabilitas dan validitas [reliability and validity]* (4th ed.). Pustaka Pelajar.

Azwar, S., & Ancok, D. (2008). *Validitas prediktif diferensial pada ujian tulis UM UGM* (Dissertation). Universitas Gadjah Mada.

Bennet, G. K., Seashore, H. G., & Wesman, A. G. (1947). *Differential aptitude tests for personnel and career assessment: Technical manual*. The Psychological Corporation.

Bennet, G. K., Seashore, H. G., & Wesman, A. G. (1952). *Differntial aptitude test: Manual* (2nd ed.). The Psychological Corporation.

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1948). The differential aptitude tests: Some comments by the authors. *Occupations: The Vocational Guidance Journal*, *27*(1), 20–22. https://doi.org/10.1002/j.2164-5892.1948.tb01449.x

Chauhan, S. S. (2009). *Advanced educational psychology* (7th ed.). Vikas Publishing House.

Corengia, A., Pita, M., Mesurado, B., & Centeno, A. (2013). Predicting academic performance and attrition in undergraduate students. *Liberabit Revista de Psicologia*, *19*(1), 101–112.

Daulay, N. (2014). Implementasi tes psikologi dalam bidang pendidikan [Implementation of psychological tests in education]. *Jurnal Tarbiyah*, *21*(2), 402–421. https://doi.org/10.30829/tar.v21i2.13

Dodonova, Y. A., & Dodonov, Y. S. (2013). Faster on easy items, more accurate on difficult ones: Cognitive ability and performance on a task of varying difficulty. *Intelligence*, *41*(1), 1–10. https://doi.org/10.1016/j.intell.2012.10.003

Doppelt, J. E., & Bennett, G. K. (1951). A longitudinal study of the differential aptitude tests. *Educational and Psychological Measurement*, *11*(2), 228–237. https://doi.org/10.1177/001316445101100206

Downey, L. A., Lomas, J., Billings, C., Hansen, K., & Stough, C. (2013). Scholastic success: Fluid intelligence, personality, and emotional intelligence. *Canadian Journal of School Psychology*, *29*(1), 40–53. https://doi.org/10.1177/0829573513505411

Dudung, A., Nurhidayat, D., & Yuliatmojo, P. (2019). The influence of engineering students numerical reasoning on the academic achievement of Computer Numerical Control (CNC) course. *Journal of Physics: Conference Series*, *1402*(4), 044053. https://doi.org/10.1088/1742-6596/1402/4/044053

Erawati, M. (2015). *Prestasi akademik siswa dalam kaitannya dengan motivasi akademik, persepsi terhadap dukungan akademik dan peran orang tua [Student academic achievement in relation to academic motivation, perceptions of academic support and the role of parents]* (Dissertation). Universitas Gadjah Mada.

Fenster, A., Markus, K. A., Wiedemann, C. F., Brackett, M. A., & Fernandez, J. (2001). Selecting tomorrow's forensic psychologists: A fresh look at some familiar predictors. *Educational and Psychological Measurement*, *61*(2), 336–348. https://doi.org/10.1177/00131640121971149

Flanagan, J. C. (1953). *FACT, interpreting test score*. Science Research Associates.

Freeman, F. S. (1976). *Theory and practice of psychological testing* (3rd ed.). Oxford & IBH Publishing Co.

Furr, R. M. (2022). *Psychometrics: An introduction* (4th ed., Vol. 5). SAGE Publications.

Gernsbacher, M. A., Soicher, R. N., & Becker-Blease, K. A. (2020). Four empirically based reasons not to administer time-limited tests. *Translational Issues in Psychological Science*, *6*(2), 175–190. https://doi.org/10.1037/tps0000232

Haciomeroglu, E. S. (2016). Object-spatial visualization and verbal cognitive styles, and their relation to cognitive abilities and mathematical performance. *Kuram ve Uygulamada Egitim Bilimleri*, *16*(3), 987–1003. https://doi.org/10.12738/estp.2016.3.0429

Hertinjung, W. S., Sulandari, S., Utari, A. P., & Hastuti, F. D. (2020). Psychological potential identification of a private vocational high school students, Indonesia: suitability of interests, aptitude, and selected majors. *Journal of Arts & Humanities*, *09*(12), 24–40. https://doi.org/https://doi.org/10.18533/jah.v9i12.2017

Hunsley, J., & Mash, E. J. (2007). Evidence-based assessment. *Annual Review of Clinical Psychology*, *3*(1), 29–51. https://doi.org/10.1146/annurev.clinpsy.3.022806.091419

IEA. (2007). *Third International Mathematics and Science Study (TIMSS 2007)* (tech. rep. No. 2007). TTIMSS International Study Center Boston College.

Kemenristekdikti. (2019). *Statistik pendidikan tinggi (Higher education statistics) 2019*. http://www.mohe.gov.my/web_statistik/

Kementerian Pendidikan dan Kebudayaan Republik Indonesia. (2014). Pemendikbud No. 49 Tahun 2014 tentang Standar Nasional Pendidikan Tinggi [Minister of Education and Culture No. 49 of 2014 concerning National Higher Education Standards [Peraturan Menteri Pendidikan dan Kebudayaan].

Krannich, M., Goetz, T., Roos, A.-L., Murayama, K., Keller, M. M., Bieg, M., & Lipnevich, A. A. (2022). Predictive validity of state versus trait challenge and boredom for career aspirations. *Learning and Instruction*, *81*, 101596. https://doi.org/10.1016/j.learninstruc.2022.101596

Kurpius, S. E. R., & Stafford, M. E. (2006). *Testing and measurement: A user-friendly guide* (Vol. 4). SAGE Publications.

Layton, W. L., & Swanson, E. O. (1958). Relationship of ninth grade differential aptitude test scores to eleventh grade test scores and high school rank. *Journal of Educational Psychology*, *49*(3), 153–155. https://doi.org/10.1037/h0046365

Lu, Y., & Sireci, S. G. (2007). Validity issues in test speededness. *Educational Measurement: Issues and Practice*, *26*(4), 29–37. https://doi.org/10.1111/j.1745-3992.2007.00106.x

Marsidi, S. R., & Hatta, A. R. (2019). Penelusuran bakat minat pada siswa SMA Di Bekasi [Talent and interest tracking for high school students in Bekasi]. *Jurnal Pengab-*

dian Masyarakat AbdiMas, *5*(2), 196–200. https://ejurnal.esaunggul.ac.id/index.php/ABD/article/view/2585/2224

Mawaddah, Ahmad, A., & Duskri, M. (2018). Gender differences of mathematical critical thinking skills of secondary school students. *Journal of Physics: Conference Series*, *1088*, 012054. https://doi.org/10.1088/1742-6596/1088/1/012054

Navarro-Soria, I., Álvarez-Teruel, J. D., Granados-Alós, L., & Lavigne-Cerván, R. (2021). Early detection of academic performance during primary education using the Spanish Primary School Aptitude Test (AEI) Battery. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.630803

Nijenhuis, J. T., Evers, A., & Mur, J. P. (2000). Validity of the differential aptitude test for the assessment of immigrant children. *Educational Psychology*, *20*(1), 99–115. https://doi.org/10.1080/014434100110416

Novalina, S. D., Sinaga, Y., Masri, R. I., & Gurusinga, D. H. (2021). Penelusuran minat bakat dan konseling psikologis dampak covid 19 bagi kehidupan sehari-hari dimasa pandemi di panti asuhan simpang tiga [Talent interest tracking and psychological counseling on the impact of covid 19 on daily life during the pandemic at the simpang tiga orphanage]. *Jurnal Guru Kita PGSD*, *6*(1), 57. https://doi.org/10.24114/jgk.v6i1.29876

Nozaleda, B. M. (2019). Validity of High School GPA and the College Aptitude Test (CAT) in predicting college academic performance. *International Journal of Scientific Research and Engineering Development*, *2*(3), 626–636. www.ijsred.com

Nugba, R. M., & Quansah, F. (2021). Standardized achievement testing, aptitude testing, and attitude testing: How similar or different are these concepts in educational assessment? *Asian Journal of Education and Social Studies*, 42–54. https://doi.org/10.9734/ajess/2021/v15i330383

Nurcahyo, F. A., Suprapto, M. H., & Boeditjahjono, J. H. (2012). Korelasi antara CFIT, tes pemahaman, dan tes perhitungan pada siswa kelas XII di Kepulauan Mentawai [Correlation between CFIT, comprehension test, and calculation test in class XII students in Mentawai Islands]. *Seminar Nasional Psikometri*, 224–232. https://publikasiilmiah.ums.ac.id/handle/11617/6414

Nurkhafifah, S., Retariandalas, R., & Megawanti, P. (2020). Pengaruh kemampuan numerik mahasiswa pendidikan matematika ditinjau dari asepek gender dan umur [The effect of numerical ability of mathematics education students in terms of gender and age]. *Matematika*, (2020), 315–322. http://www.proceeding.unindra.ac.id/index.php/DPNPMunindra/article/view/4749%0Ahttp://www.proceeding.unindra.ac.id/index.php/DPNPMunindra/article/download/4749/745

Periantalo, J. (2015). *Validitas alat ukur psikologi: Aplikasi praktis [Validity of psychological measuring instruments: Practical applications]*. Pustaka Pelajar.

Pradana, R. G., Prasetyawati, F. Y., & Mukhibun, A. (2021). Perbedaan optimisme perkuliahan daring pada mahasiswa rumpun ilmu saintek dan soshum [Differences in online lecture optimism in science and social science students]. *Jurnal Ilmiah Psikologi Candrajiwa*, *6*(1), 74–85. https://candrajiwa.psikologi.fk.uns.ac.id/index.php/candrajiwa/article/view/147

Ramist, L., Lewis, C., & McCamley-JENKINS, L. (1994). Student group differences in predicting college grades: Sex, language, and ethnic groups. *ETS Research Report Series*, *1994*(1). https://doi.org/10.1002/j.2333-8504.1994.tb01600.x

Rockinson-Szapkiw, A. J., Bray, O. R., & Spaulding, L. S. (2014). Examining the predictive validity of gre scores on doctoral education: Students' success and methodology choices in the dissertation process. *Journal of College Student Retention: Research, Theory & Practice*, *16*(2), 203–217. https://doi.org/10.2190/cs.16.2.c

Rostiana & Saraswati, K. D. H. (2018). Penelusuran Minat-Bakat untuk siswa SMA di Yogyakarta [Interest-Talent Search for high school students in Yogyakarta]. *Jurnal Bakti Masyarakat Indonesia*, *1*(1), 188–193. https://doi.org/10.24912/jbmi.v1i1.1897

Santos, J. C., & Boyon, M. C. (2020). Numerical and verbal reasoning aptitudes as predictors of STEM students' performance on limits and continuity. *11*, 14–24.

Setiawati, F. A. (2020). Aptitude test's predictive ability for academic success in psychology student. *Psychological Research and Intervention*, *3*(1), 1–12. https://doi.org/10.21831/pri.v3i1.34731

Setiawati, F. A., Izzaty, R. E., & Hidayat, V. (2018). Evaluasi karakteristik psikometrik tes bakat differensial dengan teori klasik [Evaluation of psychometric characteristics of differential aptitude tests with classical theory]. *HUMANITAS*, *15*(1), 46. https://doi.org/10.26555/humanitas.v15i1.7249

Sim, F., Thompson, L., Marryat, L., Ramparsad, N., & Wilson, P. (2019). Predictive validity of preschool screening tools for language and behavioural difficulties: A PRISMA systematic review (R. B. Gillam, Ed.). *PLOS ONE*, *14*(2), e0211409. https://doi.org/10.1371/journal.pone.0211409

Siswanto, S. (2014). Validitas sebagai alat penentuan kehandalan tes hasil belajar [Validity as a means of determining the reliability of learning outcome tests]. *Jurnal Pendidikan Akuntansi Indonesia*, *6*(1). https://doi.org/10.21831/jpai.v6i1.1795

Smout, M. F. (2019). The factor structure and predictive validity of the Kessler Psychological Distress Scale (K10) in children and adolescents. *Australian Psychologist*, *54*(2), 102–113. https://doi.org/10.1111/ap.12376

Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science?: A critical review. *American Psychologist*, *60*(9), 950–958. https://doi.org/10.1037/0003-066x.60.9.950

Sternberg, R. J., & Williams, W. M. (1997). Does the graduate record examination predict meaningful success in the graduate training of psychologists? A case study. *American Psychologist*, *52*(6), 630–641. https://doi.org/10.1037/0003-066x.52.6.630

Suramto, Harjito, P., & Kumara, A. (1996). Validitas, reliabilitas dan nilai standar dari tes pemahaman, tes penalaran dan tes berhitung [Validity, reliability and standardized scores of the comprehension test, reasoning test and arithmetic test]. *Jurnal Psikologi*, *23*(2), 74–83. https://doi.org/10.22146/jpsi.10051

Suryabrata, S. (2006). *Psikologi Pendidikan [Educational Psychology]*. PT Raja Grafindo.

Suwartono, C. (2020). Alat tes psikologi konteks Indonesia: Tantangan psikologi di era MEA [Psychological test tools in the Indonesian context: Psychology challenges in the AEC era]. *Jurnal Psikologi Ulayat*, *3*(1), 1–6. https://doi.org/10.24854/jpu37

Tierney, R. D. (2017). Fairness in educational assessment. In *Encyclopedia of educational philosophy and theory* (pp. 793–798). Springer Singapore. https://doi.org/10.1007/978-981-287-588-4_400

UNESCO International Bureau of Education and Soo Boon Ng. (2019). Exploring STEM competences for the 21st century. https://www.unesco.org

Universitas Gadjah Mada [UGM]. (2022). *Selektivitas sarjana dan sarjana terapan [Undergraduate and applied undergraduate selectivity]*. https://um.ugm.ac.id/selektivitas-sarjana-dan-sarjana-terapan/

Vera, M., & Cortes, J. A. (2021). Emotional and cognitive aptitudes and successful academic performance: Using the ecct. *International Journal of Environmental Research and Public Health*, *18*(24). https://doi.org/10.3390/ijerph182413184

Von Stumm, S., Hell, B., & Chamorro-Premuzic, T. (2011). The hungry mind: Intellectual curiosity is the third pillar of academic performance. *Perspectives on Psychological Science*, *6*(6), 574–588. https://doi.org/10.1177/1745691611421204

Wei, W., Lu, H., Zhao, H., Chen, C., Dong, Q., & Zhou, X. (2012). Gender differences in children's arithmetic performance are accounted for by gender differences in language abilities. *Psychological Science*, *23*(3), 320–330. https://doi.org/10.1177/0956797611427168

Weiner, I. B. (2013). The assessment process. In *Handbook of psychology* (pp. 3–27). John Wiley & Sons, Inc.

Woodworth, R. S. (2010). *Psychology, a study of mental life*. Project Gutenberg.

York, T. T., Gibson, C., & Rankin, S. (2015). Defining and measuring academic success. https://doi.org/10.7275/HZ5X-TX03