

The Relevance, Reliability, and Validity of the National Graduate Admission Test (N-GAT) as an Admission Tool into Graduate Programs in Ethiopia

Belay Hagos Hailu¹ and Berhanu Abera Kibret²

DOI: <https://doi.org/10.63990/ejobs.v9i1.13322>

Received: 29 December 2025; Accepted: 03 March 2026

Abstract

Addis Ababa University introduced the Graduate Admission Test (GAT) for the first time in 2020 as a screening aptitude test for candidates interested in applying for their postgraduate studies. Later, this was scaled up as a national GAT by the Ministry of Education in October 2023 and became a mandatory screening test for all applicants to graduate programs in all public and private universities. The need for the GAT arose due to concerns over the quality of graduate programs offered by various universities in Ethiopia. The objectives of the study were 1) to explore the reliability and validity of the GAT; 2) to analyze the factors associated with the test, and 3) to suggest future actions that could improve the quality of the test. Data were collected from the archives of the entire test takers who sat for the national GAT in October 2023. The entire cohort of test takers, totaling 34,520 candidates, had filled in their background information and their undergraduate GPAs, which were linked to their scores on the national GAT. Descriptive statistics, non-parametric tests (such as chi-square tests), and inferential statistics (such as t-test and analysis of variance) were used to answer the research questions. Findings of this study showed that most candidates (about 82.4%) came from four regions; that is, from Oromia (28.5%), Addis Ababa (26.2%), Amhara (17.5%), and SNNPR (10.2%), while only about 17.6% came from the other regional states. Besides, there were significantly higher numbers of male (74.7%) than female (25.3%) test takers in all locations who sat for the GAT administration in October 2023. Scores on GAT were positively associated with candidates' socioeconomic background, where those who had internet access in their homes, for instance, and those who had better undergraduate CGPA scored higher on the GAT. Furthermore, the interaction of gender and undergraduate CPA was also explored using linear regression analysis, and it was found that female candidates who had higher undergraduate CGPA were more likely to score higher on GAT as compared to their male counterparts. Furthermore, as a screening tool, the GAT was found to be highly reliable, with moderate predictive validity for candidates who joined the MSc and MA programs, but not for those who joined the PhD programs. Results were discussed and recommendations forwarded.

Keywords: Graduate Admission Test, reliability, validity, Ethiopia, postgraduate education

¹ Associate Professor of Education, College of Education and Language Studies, AAU, Email: belay.hagos@aau.edu.et

² Associate Professor, Testing Center, AAU, Email: berhanu.abera6@aau.edu.et

Introduction

Addis Ababa University (AAU) introduced the Graduate Admission Test (GAT) for the first time in October 2020 as part of its quality assurance framework. Before the introduction of GAT, AAU administered a computer-based entrance examination focusing on English language proficiency and quantitative analysis for approximately four years (from 2016-2019). During this period, the examination process was managed by the Teaching-Learning Support Center, the Graduate Program, and the Quality Assurance Offices, in collaboration with the ICT office. However, in August 2020, following consultative meetings with the then University Management Council, a consensus was reached on the need for a standardized screening tool to identify academically competent candidates for its master's and PhD programs. Followed by, the then Institute of Educational Research of Addis Ababa University and its Testing Center was then mandated to develop, administer, score, and report standardized aptitude tests that could serve as a compulsory admission test and identify candidates who could fulfill minimum academic potential to pursue their postgraduate studies at masters and PhD levels. Accordingly, a panel of psychometric experts was formed to develop, standardize, pilot test, administer, evaluate, and improve the GAT over time (Belay et al., 2021).

Moreover, the IER-Testing Center (hereafter Testing Center) established a task force and redesigned the exam by renaming the then 'Graduate Entrance Examination' as the Graduate Admission Test (GAT). The test was redesigned to include three domains of reasoning: verbal, quantitative, and analytical, which were aligned with the domains of the Graduate Record Exam (GRE). The first section of the test is dedicated to verbal reasoning. It evaluates candidates' ability to understand and analyze written passages, comprehend complex ideas, draw inferences, and evaluate arguments, including figuring out analogies, antonyms, and synonyms. The second section focuses on quantitative reasoning, assessing test takers' proficiency in understanding and interpreting numerical data, solving mathematical problems, and applying quantitative concepts. The third section emphasizes analytical reasoning, measuring candidates' capacity to analyze and evaluate logical relationships, recognize patterns, and draw conclusions based on given information. The test covers a total of 125 items including 60, 40, and 25 items for the three sections, respectively,

which have four multiple choice options, of which only one is the best answer (Belay et al., 2021).

Although the initial plan was to administer the GAT using computers, the test was conducted in a paper-and-pencil format in two rounds due to the need for more time for the development of a custom-built testing portal tailored to the unique requirements of the GAT. Consequently, about 7,496 candidates sat for the paper-and-pencil GAT, and later on, a total of 12,027 test takers applied for the test online in October 2020 (IER, 2020)³. Following the pilot testing of the testing portal (www.gat.aau.edu.et, later www.tms.aau.edu.et), the Testing Center administered the computer-based GAT for 289 applicants who missed the previous two scheduled test sessions. Then, computer-based GAT became fully operational and has been consistently administered for post-graduate applicants at AAU. After administering the GAT for two consecutive years through the system for screening applicants for AAU graduate programs, it transitioned to a national level in 2023 due to the concerns of the Ministry of Education over the quality of graduate programs in Ethiopia. That is, the Ministry of Education of the Federal Democratic Republic of Ethiopia appraised it as a good practice of screening candidates for postgraduate programs in Ethiopia and decided to scale it up at a national level for all universities, both private and public. Following this national decision by the Ministry of Education, the name was changed to National GAT (NGAT). In October 2023, AAU, in collaboration with the Ministry of Education (MoE), Ethio-telecom, and public universities, administered the NGAT nationwide through AAU's testing portal, utilizing the private network infrastructure of the MoE. In September 2024, however, the Ministry announced it would administer the NGAT for all public and private higher education institutions while AAU, as the first legally autonomous university in Ethiopia, continued administering its GAT using a dedicated testing portal.

The testing portal, www.tms.aau.edu.et, is a specialized platform designed specifically to meet the distinctive needs of the GAT. Through the portal, applicants can register for the GAT, submit required documentation, and receive updates regarding their test status. Additionally, the testing portal hosts the GAT digital interface, enabling candidates to take the exam online. It offers a secure environment for test administration, featuring functionalities

³ IER. (2020). Annual Report of the Institute of Educational Research.

for presenting test questions, recording responses, and monitoring test progress. The portal also includes features for scoring and reporting test results to candidates and relevant institutions. In a nutshell, the testing portal facilitated the GAT process, from registration to exam administration and result dissemination, while ensuring security and efficiency throughout.

The digital GAT covers 3 hours, where the timing of the three sections of sixty minutes was controlled sequentially. A candidate has to complete the first section, Verbal Reasoning, in the first sixty minutes before proceeding to the next section, Quantitative Reasoning, which requires the candidate to work on for a maximum of another sixty minutes. Similarly, candidates are allowed to work on the last section, Analytical Reasoning, during the third sixty minutes. By the end of the three-hour testing session, each test taker would be notified of their results displayed on their screening showing the raw scores, percentile scores, and standard scores (z-scores and T-scores). Besides, candidates were also able to generate their GAT performance certificate showing these results immediately after dismissal from the testing rooms (on their smartphones or PCs).

Review of the Literature

Graduate Record Examination (GRE) in Brief

Globally, general aptitude tests in higher education have been practiced to assess a test taker's academic readiness for a program that he/she applied for (ETS, 2024). Their focus is measuring reasoning abilities of the candidate rather than subject-specific knowledge (Kuncel & Hezlett, 2010). For instance, the Graduate Record Examination (GRE) is well known in this regard. The GRE, designed by the American company, Education Testing Services (ETS), and has been administered for the purpose of identifying competent graduates who could successfully complete their graduate education when admitted to the graduate programs in the USA and beyond (Powers, 2016).

The GRE followed rigorous procedure of test construction, validation, and posttest analysis and underwent psychometric refinement to make the test more relevant and fair (ETS, 2024). Similarly, Powers (2016) also asserted that frequent revisions and refinements of the GRE contributed to its measurement precision. In a meta-analysis, Kuncel and Hezlett

(2010) found that GRE was a strong predictor of academic success and argued for its fairness across diverse demographic groups. Furthermore, they asserted that predictive validity improves when GRE is combined with the effects of undergraduate CGPA, instead of relying on GRE scores as a standalone measure. A study showed that GRE had modest predictive validity GRE with GPA for graduate performance in Singaporean universities ranging from 0.15 to 0.21 (Li et al., 2014) and Dutch universities ranging from 0.17 to 0.31 (Schwager et al., 2014), which are comparable to findings reported by universities in the USA, such as Florida public universities, reporting a correlation ranging from 0.19 to .21 (Klieger et al., 2014).

Studies revealed that GRE was used for the purpose of admission of candidates into graduate programs largely in universities in the West although its predictive validity results showed inconsistent results across studies, generating debate on its relevance and fairness to diverse groups (Aciro& coauthors, 2021). Some studies also argued that GRE disproportionately favors candidates from the wealthiest backgrounds more than others, raising concerns over GRE's fairness to diverse groups despite acceptable levels of differential item functioning reports on GRE, which indicated psychometrically established fairness reports (Cheng et al., 2021). Another critique against GRE showed that the prediction power of GRE was moderate and overlooked candidates' writing skills and non-cognitive skills such as persistence (Bridgeman and Cline (2022) and limited the coverage of the construct, particularly not adequately addressing critical thinking and practical skills (Sternberg, 2020).

To sum up, the debate generally showed two narratives on the fairness and relevance of the GRE. Some argue that psychometric indicators of reliability, predictive validity, and differential item functioning showed that GRE is within the acceptable range for graduate admission decisions. Others underline concerns over the limited scope of skills tapped and favoring the wealthiest. Although GRE was used as an admission tool to graduate education in many universities in the world, admission decisions and the weight given to GRE scores vary by institution and programme (Gitomer & Wey, 2025).

As has been said, the content and purpose of the GAT are more like that of the GRE. It is designed to measure the aptitude of candidates who intend to apply to graduate programs. Verbal, quantitative, and analytical reasoning test items made up both tests similar. Besides, they both are meant to predict success at graduate programs rather than measuring past achievements of academic skills.

Reliability and Predictive validity of GAT

An initial study on GAT was carried out by Belay et al. (2021) with over 6000 test takers of the GAT and carried out item analysis, reliability, and validity indexes. Results of this study showed that GAT items were largely difficult, with only 1.6%, 20%, 69.9%, and 8% of the items being easy, moderate, difficult, and very difficult items, respectively, with moderate to low discrimination indexes (Belay et al., 2021). Furthermore, the same study revealed that the reliability of GAT is reasonably high, as estimated by five methods (N=6070): Coefficient Alpha ($\alpha = .8704$), Guttman's L2 ($\lambda_2=0.8738$), Feldt-Gilmer ($\rho=0.8728$), Feldt-Brennan ($\rho =.8723$), and Raju's Beta ($\beta =0.8704$) showed that GAT had reasonably high reliability (Belay et al, 2021).

Predictive validity was also reported in the same study by Belay et al. (2021). A total of 241 PhD students enrolled in four colleges of Addis Ababa University (i.e., Business & Economics, Natural and Computational Sciences, Social Sciences and Addis Ababa Institute of Technology) showed the very low relationship between GAT and first semester GPA was not statistically significant ($r=.066$, $N=241$), indicating poor predictive power of GAT for PhD programs (Belay, et al., 2021). On the other hand, modest ($r=0.16$ to $r=.39$) predictive validity index of GAT were found in the same four colleges of Addis Ababa University as correlated with the first semester GPA of students ($N=593$) enrolled for their second degrees (MA/ MSc programs) (Belay, et al., 2021). The predictive validity indexes vary for the four colleges: College of Business & Economics ($r=.39$), College of Natural and Computational Sciences ($r=.217$), College of Social Sciences ($r=.223$), and Addis Ababa Institute of Technology ($r=.164$) (Belay et al., 2021). Finally, the study concluded that GAT has a relatively high reliability index, with evidence of predictive validity for master's programs but not for PhD programs, which requires further studies on admission policies for terminal

degree applicants. Even then, the study underlined that GAT fell within the acceptable ranges of item characteristics.

Statement of the Problem

Universities and the Ministry of Education have been concerned about the quality of university education in general and graduate programs in particular. It is expected that graduate education is only for those who are highly competent candidates who could be engaged in scientific endeavors with higher analytical and problem-solving skills. The graduate admission policies varied by universities, and gaps and issues of quality emerged, which necessitated the emergence of a national-level graduate admission test. However, there have been fewer studies on the reliability, validity, and relevance of the graduate admission test (GAT); the issue was not studied. Furthermore, as a member of the task force involved in the development and implementation of the digital GAT since its inception, the researchers have observed varying opinions and concerns expressed by both test takers and administrators regarding the usability of the test and the use of cut-off scores or weighted composite index. The objectives of the study were, therefore, 1) to explore the reliability and validity of the GAT; 2) to analyze the factors associated with the test, and 3) to suggest future actions that could improve the quality of the test. More specifically, the study attempts to address the following specific research questions:

- 1) Are gender and other demographic characteristics associated with national graduate admission test (GAT) performances?
- 2) Does the perceived socio-economic status of national GAT candidates have a relationship with the national graduate admission test (GAT)?
- 3) Is the national GAT a reliable and valid measure relevant for the purpose of screening applicants for graduate admission?
- 4) To what extent is undergraduate cumulative grade point average (CGPA) related to scores on the national graduate admission test (GAT)?
- 5) What are the rationale and optional cutoff scores for passing the national graduate admission test (GAT)?
- 6) What lessons can be drawn for better admission decisions into graduate programs?

Methods

Data were collected from the Testing Center archives of the Institute of Educational Research of Addis Ababa University. The entire number of test-takers who sat for the national GAT in October 2023 was 34,520, with 25.3% female and 74.7% male. The test takers had filled in their background information as well as their results on their undergraduate GPAs, which were mandatory before the test began. As a result, their unique identification number was used to connect their background information with their scores on the national GAT. Descriptive statistics, non-parametric tests (such as chi-square tests), and inferential statistics (such as t-test and analysis of variance) were used to analyze the data and answer the research questions.

Results

Background characteristics of test takers

A total of 34,520 candidates sat for the October 2023 national GAT administration. About 25.3% and 76.5% female and male candidates, respectively, sat for the October 2023 GAT administration. When the pass rates were released, about 26% of the female and 74% of the male test takers got the pass marks, with relatively similar proportions. Even then, the participation of female candidates remained quite low.

The majority of the GAT candidates (82.4%) completed their secondary school education from Oromia (28.5%), Addis Ababa (26.2%), Amhara (17.5%), and SNNPR (10.2%) while only 17.6% of the candidates were from Afar (1%), Benshangul-Gumuz (.9%), Dire Dawa (1.2%), Gambella (.4%), Harari (.8%), Sidama (2.5%), Somalie (2.9%), Tigray (5.7%), and others such as those from Somaliland, Puntland, South Sudan (2.1%). More female candidates than the national average (i.e., 25.3%) were recorded among those applicants from Addis Ababa (41.5%), Puntland (38.6%), Harari (31%), and Dire Dawa (28.8%).

Table 1

Background characteristics of GAT takers

Location of completion of secondary school	N	Female	Male	Total	Chi-square test	
1) Addis Ababa	9,043	41.50%	58.50%	26.20%		
2) Afar	356	17.10%	82.90%	1.00%		
3) Amhara	6,052	21.00%	79.00%	17.50%		
4) Benshangul-Gumuz	317	19.60%	80.40%	0.90%		
5) Dire Dawa	424	28.80%	71.20%	1.20%		
6) Gambela	135	6.70%	93.30%	0.40%		
7) Harari	277	31.00%	69.00%	0.80%	$X^2=1821$, df=12, p<.05	
8) Oromia	9,825	19.20%	80.80%	28.50%		
9) Sidama	860	22.90%	77.10%	2.50%		
10) Somalie	1,008	12.90%	87.10%	2.90%		
11) SNNPR	3,527	17.20%	82.80%	10.20%		
12) Tigray	1,980	23.50%	76.50%	5.70%		
13) Other (Somaliland, Puntland, South Sudan)	716	12.60%	87.40%	2.10%		
Total	34,520	25.30%	74.70%	100.00%		
Country	N	Female	Male	Total		
a) Ethiopia	33,762	25.60%	74.40%	97.80%		
b) Puntland	70	38.60%	61.40%	0.20%		$X^2=96.77$, df=3, p<.05
c) Somaliland	649	9.20%	90.80%	1.90%		
d) South Sudan	39	23.10%	76.90%	0.10%		
Total	34,520	25.30%	74.70%	100.00%		

Only very few candidates (.58%) were applicants with disabilities, largely those with motor and visual impairments. There were significantly higher number of males (79%) with disabilities than females (21%) who sat for the national GAT as compared to the proportion for candidates without disabilities ($X^2=1.99$, df=1, p>.05), showing female with disabilities disproportionately low number of applicants into graduate programs.

Table 2

GAT candidates by gender and disability type

Do you have any disability?	N	Female	Male	Total	Chi-square test
No	34,320	41.50%	58.50%	26.20%	$X^2=1.99$, df=1, p>.05
Yes	200	21.00%	79.00%	60.00%	
Total	34,520	25.30%	74.70%	100.00%	
What type of disability?					
Hearing impairment	19	21.10%	78.90%	0.06%	$X^2=3.60$, df=4, p>.05
Visual Impairment	50	18.00%	82.00%	0.14%	
Physical /Motor Impairment	97	19.60%	80.40%	0.28%	
Other disabilities	34	29.40%	70.60%	0.10%	
No disability	34,320	25.30%	74.70%	99.40%	
Total	34,520	25.30%	74.70%	100.00%	

As shown in Table 3, the majority of the candidates (76%) reported that their family's socioeconomic status was moderate or the same as that of most people in their community, while 16.9% reported it as worse than that of most people in their communities. On the other hand, only very few of them (7.1%) reported that their family's socioeconomic status was better than that of most people in their community. The proportion of males and females in these three categories was not the same since relatively fewer females (14.7%) endorsed that their families' socio-economic status was worse than most people ($X^2=160.3$, $df=2$, $p<.05$). In other words, the female candidates were from relatively better socio-economic backgrounds. Slightly more than half of the candidates (57.4%) said they had internet at their homes. There were significantly higher numbers of female candidates who reported they had internet at their homes than those who said they didn't have ($X^2=408.69$, $df=1$, $p<.05$), supporting the earlier argument that female candidates were from better socioeconomic backgrounds. In terms of employment after graduation, the majority (87.7%) reported that they were employed. However, the unemployment proportion was higher among female candidates ($X^2=30.21$, $df=1$, $p<.05$).

Table 3

Gender of candidates by perceived socioeconomic status, internet access at home, and employment status

As compared to most people in your community, how do you rate your family's socioeconomic status?	N	Female	Male	Total	Chi-square test
Better than most people	2,449	26.80%	73.20%	7.10%	$X^2=160.3,$ df=2, p<.05
Same as most people	26,214	26.00%	74.00%	76.00%	
Worse than most people	5,847	14.70%	85.30%	16.90%	
Total	34,510	25.30%	74.70%	100.00%	
Do you have internet access at home?					
No	14,708	19.80%	80.20%	42.60%	$X^2=408.69,$ df=1, p<.05
Yes	19,812	29.40%	70.60%	57.40%	
Total	34,520	25.30%	74.70%	100.00%	
Were you employed after graduation?					
No	4,253	28.80%	71.20%	12.30%	$X^2=30.21,$ df=1, p<.05
Yes	30,267	24.80%	75.20%	87.70%	
Total	34,520	25.30%	74.70%	100.00%	

Factors associated with GAT

In this section, factors, such as gender, socio-economic status, access to the internet, disability, and location were explored to determine if they had associations with GAT results. To begin with, let's see the descriptive statistics of the national GAT in T-scores⁴.

⁴ Notes: T-scores are standard scores that are transformed with a predetermined mean of 50 and a standard deviation of 10 (i.e., $50+10Z$) just to make the scores of different groups and administrations comparable. Z-score is the basic standard score generated from the raw scores, previously established mean and standard deviation (which was a normative reference data); that is, $Z=(X-\text{Mean}/SD)$.

Table 4

Descriptive Statistics in T-scores on the National GAT for the October 2023 Administration at Addis Ababa University

<i>GAT</i>	<i>N</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>	<i>Skewness</i>	<i>Kurtosis</i>
Verbal	34,515	22.78	89.79	51.63	10.25	0.494	-0.036
Quantitative	34,442	23.17	88.43	45.29	7.96	0.717	0.886
Analytical	34,451	24.82	76.58	48.02	7.97	0.457	-0.006
GAT Overall	34,515	14.37	90.17	48.27	9.81	0.609	0.200

The national GAT T-score ranged as low as 14.37 and as high as 90.17. In other words, it ranged between a Z-score of -3.56 and +4.02. The score distributions were also within the range of the normal distribution as the measures of skewness and kurtosis indicated. The candidates in general performed slightly higher on verbal and analytical sub-tests than in quantitative sub-test of the national GAT.

Table 5

Mean Scores for National GAT for the October 2023 Administration and their undergraduate CGPA by Gender

<i>Gender</i>		<i>N</i>	<i>M</i>	<i>SD</i>	<i>T</i>	<i>df</i>	<i>P</i>	<i>Cohen's d</i>
Verbal	Female	8742	51.45	10.79	-1.95	34513	.052	
	Male	25773	51.70	10.06				
Quantitative	Female	8730	44.46	7.58	-11.29	34440	.001*	-.140
	Male	25712	45.57	8.06				
Analytical	Female	8729	48.02	8.44	0.05	34449	.959	
	Male	25722	48.02	7.81				
GAT-Total	Female	8742	47.84	10.18	-4.83	34513	.001*	-.060
	Male	25773	48.42	9.67				
Undergraduate CGPA	Female	8559	3.11	.47	-15.59	33926	.001*	-.195
	Male	25,369	3.20	.45				

Note. *M* = Mean; *SD* = Standard Deviation. An asterisk indicates $p < .05$.

Raw scores were converted to T-scores to make them comparable across groups ($T\text{-score} = 50 + 10Z$).

One of the research questions in this study focused on whether there was a gender difference in the overall GAT and sub-components. An independent-samples t-test revealed a

significant difference on the overall performance of the national GAT between females ($M = 47.84, SD = 10.18$) and males ($M = 48.42, SD = 9.67$), $t(34513) = -4.83, p < .001$, indicating lower performance by female candidates with low effect size ($d = -.06$). Besides, there was a statistically significant difference on quantitative reasoning sub-test of the national GAT between females ($M=44.46, SD=7.58$) and males ($M=45.57, SD=8.06$), $t(34440) = -11.29$, indicating lesser performance by female candidates with low effect size ($d=-.140$).

On the other hand, there was no statistically significant difference on the verbal reasoning sub-test of the national GAT between females ($M = 51.45, SD = 10.79$) and males ($M = 51.70, SD = 10.06$), $t(34513) = -1.95, p = .052$. Similarly, there was no statistically significant difference on the analytical reasoning sub-test of the national GAT between females ($M = 48.02, SD = 8.44$) and males ($M = 48.02, SD = 7.81$), $t(34449) = .959, p = .052$, indicating similar performance.

In a related analysis, we explored whether female and male candidates were different in their overall undergraduate academic records. This study revealed that there was a statistically significant difference in the candidates' undergraduate cumulative grade point average (C-GPA) between females ($M=3.11, SD=.47$) and males ($M=3.20, SD=.45$), $t(33926) = -15.59$, showing lesser performance by female candidates with a low effect size ($d=-.195$).

Table 6.

Regression Predicting National GAT Performances

Predictor	<i>B</i>	<i>SE</i>	95% <i>CI</i>	β	<i>p</i>
Gender (Male-dummy)	.082	.121	[-.155, .318]	.004	.498
CGPA (centered)	5.388	.218	[4.961, 5.815]	.254	.001
Interaction (Gender X CGPA)	-1.215	.255	[-1.715, -.716]	-.049	.001
Constant	48.322	.104	[48.118, 48.527]		.000

Note. *b* = unstandardized regression coefficient; *SE* = standard error; *CI* = confidence interval; β = standardized coefficient. $p < .05$.

Linear regression analysis was conducted to predict national GAT scores based on gender and undergraduate academic performance as measured by the CGPA. Results showed that gender alone was not a statistically significant predictor of national GAT scores ($b=.082$,

$p=.498$), academic records during undergraduate studies, or the CGPA, which was a significant positive predictor ($b=5.388, p=.001$), showing that as the CGPA increases by one unit, the national GAT score increases by 5.388 points. Furthermore, the interaction between gender and CGPA was statistically significant ($b = -1.215, p = .001$). For female candidates a one-unit increase in CGPA was associated with 5.388 points increase in the national GAT; however, the one-point increase in CGPA in males was associated with 4.173-point increase in the national GAT (i.e., 1.215 points less).

Table 7

Mean Scores for National GAT for the October 2023 Administration by Internet Access at Home

Do you have internet access at home?		N	Mean	SD	t	df	Sig.	Cohen's d
Verbal	No	14,523	50.04	9.07	-25.14	34,179	.001	-.275
	Yes	19,658	52.81	10.79				
Quantitative	No	14,480	44.80	7.67	-10.05	34,107	.001	-.110
	Yes	19,629	45.68	8.15				
Analytical	No	14,485	47.05	7.45	-19.74	34,117	.001	-.238
	Yes	19,634	48.77	8.28				
GAT-Total	No	14,523	46.80	8.77	-24.41	34,179	.001	-.267
	Yes	19,658	49.39	10.33				

Note: T-Score has a mean of 50 and SD of 10 [i.e., $T=50+10Z$]; $Z\text{-score}=(X\text{-Mean})/SD$.

Whether candidates who had or did not have internet access at home was associated with their performances on the overall as well as the sub-components of GAT. An independent-samples t-test was run and a significant difference was found on the overall performance of the national GAT between females ($M = 46.80, SD = 8.77$) and males ($M = 49.39, SD = 10.33$), $t(34179) = -24.41, p < .001$, indicating lower performance by candidates who had no internet access at their homes, with small effect size ($d = -.267$). Similar results of higher national results were recorded on the sub-tests (verbal, quantitative, and analytical reasoning sub-tests) for candidates who had internet access in their homes.

Table 8

One-way Analysis of Variance of National GAT Results of Candidates by their Perceived Socio-Economic Status (SES)

As compared to most people in your community how do you rate your family's socioeconomic status?						
Responses	<i>N</i>	<i>M</i>	<i>SD</i>	<i>F</i>	η^2	
Worse than most people	2,449	46.74	9.51	161.61	.009	
Same as most people	26,211	48.81	9.79			
Better than most people	5,845	46.54	9.75			
Total	34,505	48.28	9.81			

Note. *M* = Mean; *SD* = Standard Deviation. $F(2, 34502) = 161.61, p = .001, \eta^2 = .009$.

Table 9

One-way ANOVA comparing perceived Socio-Economic Status on National GAT Results

Source	Sum of Squares	Df	Mean Square	<i>F</i>	Sig.
Between Groups	57975.2	2	28987.6	161.6	0.000
Within Groups	6188370.2	34502	179.4		
Total	6246345.4	34504			

One-way analysis of variance revealed a statistically significant difference in national GAT scores across three conditions of perceived socio-economic status, $F(2, 34502) = 161.61, p = .001$, with low effect size, $\eta^2 = .009$. Post-hoc comparisons of results using Scheffé SHD showed that those who claimed their perceived SES was the same as most people in their communities, ($M = 48.81$) scored significantly higher on the national GAT than those candidates who claimed their SES was worse ($M = 46.74$), with a mean difference of about 2.84, $p = .001$, 95% CI [2.15, 3.53]. Likewise, those who claimed their SES was the same as most people performed higher on national GAT ($M = 48.25$) than those who claimed their SES was better ($M = 45.14$), with a mean difference of about 3.11, $p = .001$, 95% CI [2.64, 3.58], $p = .001$. This result showed that those who were perceived as candidates from better socio-economic backgrounds were not any better in the GAT test performances than the other humble groups.

Table 10

Post-hoc Comparisons (Scheffee SHD) of Perceived SES on National GAT Results

As compared to most people in your community, how do you rate your family's socioeconomic status?					95% Confidence Interval	
I	J	Mean Difference (I-J)	SE	p	Lower Bound	Upper Bound
Worse than most people	Same as most people	-2.84*	.283	0.00	-3.53	-2.15
	Better than most people	0.27	0.32	0.70	-0.52	1.06
Same as most people	Worse than most people	2.84*	0.28	0.00	2.15	3.53
	Better than most people	3.11*	0.19	0.00	2.64	3.58
Better than most people	Worse than most people	-0.27	0.32	0.70	-1.06	0.52
	Same as most people	-3.11*	0.19	0.00	-3.58	-2.64

*. The mean difference is significant at the 0.05 level.

Table 11

Mean Differences in National GAT Results of Candidates by Disability Conditions

Disability Conditions		N	M	SD	t	Df	p	Cohen's d
Are you a person with disability?	No	34,315	48.30	9.80	7.82	34,513	.001	.554
	Yes	200	42.87	9.09				

Notes: Disability Type included: Visual, Hearing, and Motor Disabilities

Candidates with disabilities also participated in the October 2023 GAT administration, and their performance was compared with the large majority of candidates without disabilities. An independent samples t-test result revealed that there was a statistically significant difference on the national GAT performances between candidates without disability ($M=48.30, SD=9.80$) and those with disability ($M=42.87, SD=9.09$), $t(34513)=7.82$, $p=.001$, 95% CI [4.07, 6.79], with moderate effect size [$d=.554$], indicating persons with disabilities scored lower on the GAT.

Table 12

One-way Analysis of Variance of National GAT Results of Candidates by Location of their Country

<i>Location</i>	<i>N</i>	<i>Mean</i>	<i>SD</i>	<i>F-value</i>	<i>Cohen's d</i>
Ethiopia	33,757	48.31	9.76	14.2	.001
Punt Land	70	43.20	5.81		
Somali Land	649	47.54	12.10		
South Sudan	39	41.28	10.00		
Total	34,515	48.27	9.81		

Note: The Staff of the Testing Center of Addis Ababa University travelled outside of Ethiopia and administered the national GAT in these locations on October 2023.

Note. *M* = Mean; *SD* = Standard Deviation. $F(3,34511) = 14.21$; $p < .05$; $\eta^2 = .001$.

One way analysis of variance showed a statistically significant difference in national GAT scores across candidates from four locations, $F(3, 34511) = 14.21$, $p = .001$, with very low effect size, $\eta^2 = .001$, showing candidates from Ethiopia had significantly higher GAT scores than those from Puntland and South Sudan; likewise, candidates from Somaliland had significantly higher GAT scores than those from Puntland and South Sudan.

Psychometric Properties of National GAT

In this section, we will see how the three subtests of the GAT are correlated with the total GAT results and with the candidates' undergraduate academic performances as measured by their CGPAs.

The inter-correlations of the subtests are moderate (i.e., $r = .366$ to $r = .573$), which are within acceptable ranges. Higher correlations were observed between the sub-tests and the overall GAT score (i.e., $r = .693$ to $r = .890$), indicating that each component contributed significantly and meaningfully to the overall scores.

The relationship between CGPA and GAT was also found to be positive though the strength of the relationship was not that strong (i.e., $r = .165$ to $r = .212$). In general, those candidates with higher CGPA are likely to get higher GAT scores and vice versa for those who scored lower on the CGPA.

Table 13

Inter-correlations among GAT sub-tests, GAT, and CGPA

<i>GAT & CGPA</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
A. Undergraduate CGPA	1				
B. GAT-Total	.212**	1			
C. Verbal	.179**	.890**	1		
D. Quantitative	.165**	.694**	.366**	1	
E. Analytical	.174**	.793**	.573**	.475**	1
N	33,860	34,451	34,451	34,423	34,451

** . Correlation is significant at the 0.01 level (2-tailed).

Reliability of GAT

The reliability of the national GAT consisted of 125 items was also calculated as measured by Cronbach's Alpha, which is a measure of internal consistency, and was found to be highly reliable ($\alpha = .87$) indicating that if candidates took the same test repeatedly the probability that they would get similar results would be about 87% of the cases which is more than the expected threshold of .70. Hence, the national GAT had acceptable level of reliability index.

Options for possible cutoff scores and estimations of pass rates

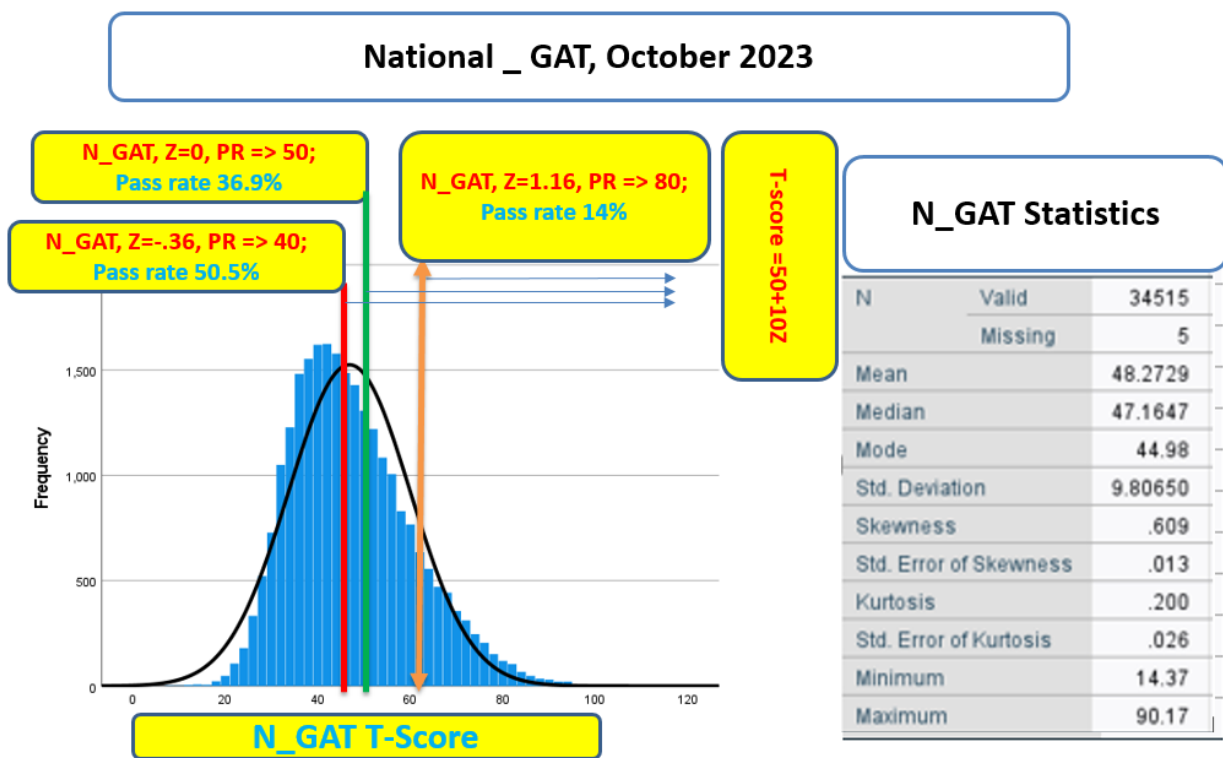
The score distribution of the national GAT of 34,515 valid cases, which was administered in October 2023, was normally distributed as the measures of skewness and kurtosis were within range. There were three possible scenarios suggested for determining the cutoff scores for passing the GAT, ranging from a relatively lenient to more stringent cutoff score.

Option 1. Cutoff score set at 40 percentile ranks, which is equivalent to Z-score = -.36 or T-score = 46.44. This cutoff score is relatively lenient but acceptable as the bottom 49.5% of the candidates would be excluded from applying for the graduate program and, even then, not all with such minimum score (i.e., a percentile score of 40 and above) would be admitted to

the graduate program as the departments had their own additional department relevant screening tools. If this was considered as cutoff score by the university management they could get as many as 50.5% of the candidates as eligible candidates into their graduate programs although departments may reject some of these candidates when failing to meet department level requirements.

Figure 1

Normal Distribution Statistics on National GAT (October 2023), T-score, Z-score, and Percentile Rank (PR) with three possible cutoff scores and estimated pass rates



Option 2. Cutoff score set at 50 percentile rank, which is equivalent to Z-score = 0.0 or T-score = 50.0. This cutoff score is moderately strict decision as it would exclude as many as 63.1% of the candidates from the application processes into the graduate programs; only 36.9% of the candidates would apply into the existing graduate programs. Even the, not all

these would be admitted into the graduate programs as departments have further screening tools of admission.

Option 3. Cutoff score set at 80 percentile rank, which is equivalent to Z-score = 1.16 or T-score = 61.74. This cutoff score is more stringent decision as it would exclude as many as 85.58% of the candidates from the application processes into the graduate programs; only about 14% of the candidates would apply to the existing graduate programs. This decision might deplete the number of applicants into the graduate programs as the cutoff score is set too high to make it practical. The National GAT, which was administered in October 2023, had a cutoff score set at the 80th percentile rank, and many graduate programs did not receive enough applicants.

Discussion

Graduate admission policies have not been uniformed all across universities in the world. However, many universities rely on standardized aptitude tests that can predict the successful completion of candidates when they join their graduate studies. One of the prominent standardized aptitude tests is the Graduate Record Examination (GRE), designed and implemented by the Education Testing Services (ETS), an American private company, which claims the GRE evolved over time as a reliable, valid, and relatively fair construct used for graduate admission (ETS, 2024). The Graduate Admission Test (GAT) is similar to the GRE in that it measures an aptitude test composed of verbal, quantitative, and analytical reasoning test items with the intention of selecting competent candidates for graduate program admission.

One of the findings of this study showed that there was a statistically significant difference in the candidates' undergraduate cumulative grade point average (C-GPA) between females and males, showing lesser performance by female candidates. Similarly, male candidates scored higher on the national GAT than female candidates. This finding is in line with the study by Miller and Stassun (2014), which found that women from low socio-economic backgrounds scored lower on the GRE. These findings are in line with the results found by Walpoleet al. (2002) where candidates from low socio-economic backgrounds who

sat for the GRE scored lower. Even then, this finding was not attributed to a lack of potential of the candidates but to the limited opportunities for preparation and cultural barriers facing women from disadvantaged backgrounds (Miller & Stassun, 2014). The access to the internet at home was positively associated with GAT scores in this study although this was not in line with the findings by Beuermann et al. (2019), who found that internet access did not contribute to candidates test scores.

The measure of perceived socio-economic status showed that those who claimed their families were neither poor nor rich, or those who said their families were average (i.e., the same as most people in their communities), scored higher on the GAT than the other two groups (i.e., those who claimed they were better or worse than most people in their communities). Besides, those who had access to the internet in their homes had higher GAT scores than those who had no access to internet. This shows that the objective measure of the presence of internet access was associated more positively with the GAT score than the subjective measure of socio-economic status. Besides, persons with disabilities scored lower than those without disabilities demanding some kind of affirmative action to encourage their participation without compromising merit-based selection.

The main and interaction effects of gender and undergraduate GPA indicated that the prediction of performance on GAT was not the same. The predicted model is $\hat{Y} = 48.322 + .082A + 5.388B - 1.215C$, where \hat{Y} = predicted GAT; A = Gender (male-dummy); B = Undergraduate CGPA; C = Interaction of Gender by GPA. Although GPA is a predictor of GAT results, the effect is not the same for females and males; the positive influence is less pronounced for males than for females. In other words, as revealed by the regression coefficients, for every 1-point increase in GPA, the national GAT score increases by about 5.39 points although this is reduced by 1.22 points for males as compared to females.

What should be the next course of action that could improve admission policies?

Cutoff scores for passing the GAT can be logically suggested by experts in the field, but the top management of an institution, such as universities or the Ministry of Education can choose feasible scenarios. The three scenarios suggested a percentile rank of 40, 50, or 80, with a possibility of about 50.5%, 39.6%, and 14% pass rates, respectively, as estimated

from the actual data in this study. These cutoff scores do not necessarily mean acceptance rates into graduate programs. It only means that these candidates can apply to graduate programs of their preferences, and those scoring below the specified percentile rank will not be allowed to submit their application. The first scenario that suggests a percentile rank of 40, for instance, is in line with the practice of some of the universities in the USA, such as the University of Florida that use TOEFL and IELTS as its standardized admission tools (Klieger et al., 2014), with an acceptance rate of 46%. Other universities use more stringent criteria of admission with lower acceptance rates. In the Ethiopian context, considering 40 percentile rank as a cutoff score on the GAT means there will be about half of the candidates who could possibly apply into graduate programs, while the remaining half will not have access to application into graduate programs. If the criteria are more stringent, and the cutoff score is increased to the 50th percentile rank, then there will be about 40% of the test takers who could possibly apply for graduate programs while the majority, or about 60% , will not have access to submit their application to graduate programs. If the cutoff score is very high, such as 80 percentile rank, as it was decided in October 2023 by the Ministry of Education, then there will be only about 14% of the candidates who will be able to apply to graduate programs, and the remaining 86% will not be able to apply for graduate studies. In this case, all those who scored above the cutoff score are not granted admission by the department, as they have their subject-specific screening tests as well. As a result, pushing the cutoff score too high is not recommended as the purpose of the NGAT is not the selection of gifted candidates; it rather focuses on leaving out those less capable candidates, and as a result, are unlikely to successfully complete graduate studies upon joining the programs. What seems more feasible and logically acceptable, as well as shared by many Western universities, is to set the cutoff scores either at 40 or 50 percentile ranks and refine the screening of candidates at the department level.

One of the recommendations of this study is introducing an additional screening tool, such as a measure of analytical reasoning through a written test, particularly for PhD candidates. Besides, transforming the GAT tests into computer adaptive testing (CAT) where each candidate's ability level is determined based on his/her responses to each test item. CAT requires item-level testing and storing the tested items in the item bank for later use.

Furthermore, test security should be a priority to ensure the CAT measures what it intends to measure.

Declaration of Conflicting Interests

No potential conflict of interest was reported by the authors.

Funding information

The authors received no financial support for this research.

Acknowledgements

The authors thank the former Institute of Educational Research for providing access to the data on Graduate Admission Test (GAT)..

References

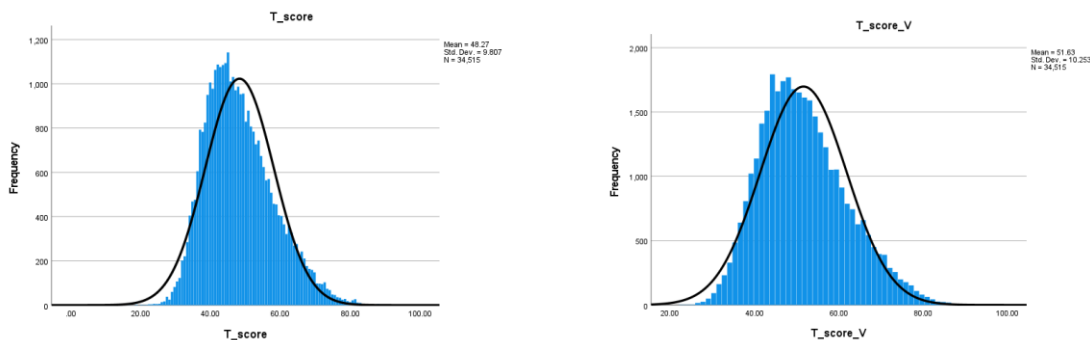
- Aciro, R., & coauthors. (2021). Test scores, socioeconomic status, and admissions: International perspectives. *Education Quarterly Reviews*, 8(4), 112–130. <https://files.eric.ed.gov/fulltext/EJ1292211.pdf>
- Belay Hagos Hailu, Berhanu Abera, Daniel Desta, & Girma Lemma. (2021). General Admission Test (GAT) at Addis Ababa University: Initial Results on its Reliability and Validity. Testing Center at the Institute of Educational Research, Addis Ababa University, unpublished manuscript.
- Beuermann, D. W., Cristia, J. P., & Cueto, S. (2019). The effects of internet access on student learning: Evidence from Peru. *Economics of Education Review*, 70, 1–19. <https://doi.org/10.1016/j.econedurev.2019.03.001>
- Bridgeman, B., & Cline, F. (2022). Can the GRE predict valued outcomes? Dropout and writing skill. *PLOS ONE*, 17(6), e0268738. <https://doi.org/10.1371/journal.pone.0268738>
- Cheng, L., Klinger, D., & Zheng, Y. (2021). Test preparation, test performance, and the role of coaching for GRE takers. *Assessment in Education: Principles, Policy & Practice*, 28(4), 421–439. <https://doi.org/10.1080/0969594X.2020.1851847>

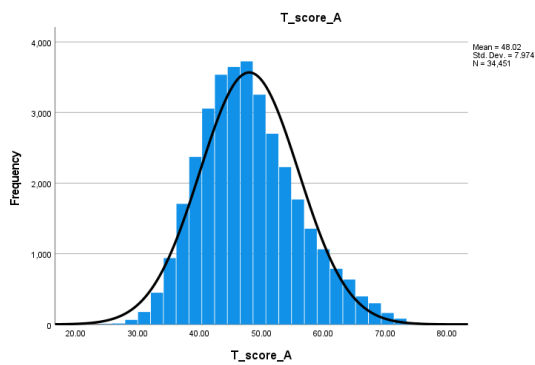
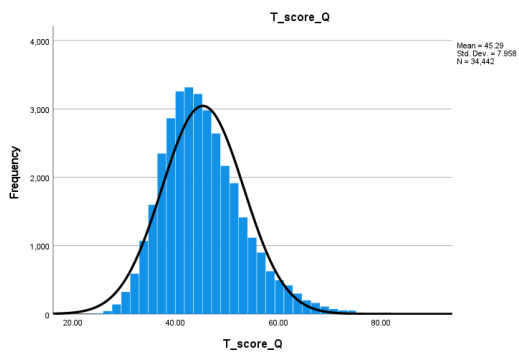
- Educational Testing Service. (2024). GRE: Graduate Record Examinations – Technical Manual. Princeton, NJ: ETS.
- Gitomer, D. H., & Wey, A. (2025). The impact of the GRE on admissions decisions: An exploratory study. *Applied Measurement in Education*. Advance online publication. <https://doi.org/10.1080/08957347.2025.2514441>
- IER. (2020). Annual Report of the Institute of Educational Research. Unpublished, Addis Ababa University.
- Klieger, D. M., Cline, F. A., Holtzman, S. L., Minsky, J. L., & Lorenz, F. (2014). New perspectives on the validity of the GRE® General Test for predicting graduate school grades (GRE Board Research Report No. 14-03). Princeton, NJ: Educational Testing Service. 10.1002/ets2.12026
- Kuncel, N. R., & Hezlett, S. A. (2010). Fact and fiction in cognitive ability testing for admissions and hiring decisions. *Current Directions in Psychological Science, 19*(6), 339–345. <https://doi.org/10.1177/0963721410389459>
- Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: Implications for graduate student selection and performance. *Psychological Bulletin, 127*(1), 162–181. <https://doi.org/10.1037/0033-2909.127.1.162>
- Liu, J., Klieger, G. W., Bochenek, J., Holtzman, S., & Xu, J. (2016). *Exploring the score profiles of international students intending to pursue a graduate degree in the United States* (Research Report No. RR-16-10). Educational Testing Service. <https://doi.org/10.1002/ets2.12097>
- Miller, C., & Stassun, K. (2014). A test that fails. *Nature, 510*(7504), 303–304. <https://doi.org/10.1038/nj7504-303a>
- Moneta, G. B., & van Helvert, J. (2007). Gender differences in GRE scores of students from developing countries. *Higher Education, 54*(3), 335–345. <https://doi.org/10.1007/s10734-006-9001-2>
- Ou Lydia Liu, David M. Klieger, Jennifer L. Bochenek, Steven L. Holtzman, & Jun Xu (2014). An Investigation of the Use and Predictive Validity of Scores From the GRE® revised General Test in a Singaporean University. ETS Research Report.

- Powers, D. E. (2016). Understanding the GRE revised General Test. *ETS Research Report Series, 2016(1)*, 1–16. <https://doi.org/10.1002/ets2.12092>
- Schwager, I. T., Hülshager, U. R., Bridgeman, B., & Lang, J. W. (2015). Graduate Student Selection: Graduate Record Examination, Socioeconomic Status, and College GPA as Predictors of Program Completion. *Journal of Applied Psychology, 100(6)*, 1855–1867. <https://doi.org/10.1037/apl0000021>
- Schwager, I., Hülshager, U., & Lang, J. (2014). Validity of the GRE® General Test and ETS® PPI in a European context. (Unpublished manuscript).
- Schwager, I., Hülshager, U., Lang, J., & Bridgeman, B. (2014). Improving fairness in graduate student selection: GRE, socioeconomic status, and undergraduate grade point average as predictors of study success in a Western European university. (Unpublished manuscript).
- Stemper, S., & Jain, R. (2024). *3G Internet and human capital development*. University of Auckland Business School.
- Sternberg, R. J. (2020). The road to measuring critical thinking: The case of the GRE. *Psychology in the Schools, 57(10)*, 1563–1572. <https://doi.org/10.1002/pits.22422>
- Walpole, M., Burton, N. W., Kanyi, K., & Jackenthal, A. (2002). *Selecting for success: Graduate admissions and associate degree students* (GRE Board Report No. 98-08P). Educational Testing Service.

Annex

Normal distributions of NGAT and its three components





Notes: Normal distribution of GAT scores; T-score is for the overall GAT; T_Score_V = verbal reasoning; T_Score_Q=quantitative reasoning; T_Score_A=analytical reasoning

		Statistics			
		T_score	T_score_V	T_score_Q	T_score_A
N	Valid	34515	34515	34442	34451
	Missing	5	5	78	69
Mean		48.2729	51.6349	45.2920	48.0205
Median		47.1647	50.5942	44.3386	47.5983
Mode		44.98	44.27	42.57	47.60
Std. Deviation		9.80650	10.25263	7.95832	7.97394
Skewness		.609	.494	.717	.457
Std. Error of Skewness		.013	.013	.013	.013
Kurtosis		.200	-.036	.886	-.006
Std. Error of Kurtosis		.026	.026	.026	.026
Minimum		14.37	22.78	23.17	24.82
Maximum		90.17	89.79	88.43	76.58

Notes: Descriptive Statistics of GAT