GENERAL (APTITUDE) MATH EXAMS AND SUCCESS IN MATHEMATICAL CLASSES AT UNIVERSITY: A CROSS-COUNTRY COMPARISON

Genady Grabarnik[1], Luiza Kim-Tyan[2] and Serge Yaskolko[3]
[1]Dr., Department of Math & CS, St. John's University, Queens, NY, USA
[2]Dr., Department of Mathematics, NUST MISIS, Moscow, Russia
[3]Dr., AD&D, College Board, Pittsburgh, PA, USA

ABSTRACT

The goal of the paper is to understand to what degree the student success depends on teaching practices and materials, and to what degree it is due to differences in background, including prior knowledge, skills, preparation, abilities, etc. We analyzed the existing research on outcomes of the SAT and ACT in the United States and the Unified Country Exam (UCE/ЕГЭ) in the Russian Federation, since those exams usually considered as a good indicator of skills, preparation, and abilities. In the US, exams concentrate on identifying success in college studies. In the RF, exams concentrate on identifying gifted students. According to our analysis, the tests are based on different goals which they respectively satisfy.

KEYWORDS

College Readiness, Assessment, Aptitude Tests, Predicting Success in College STEM Courses

1. INTRODUCTION

In our daily work, we aim to help STEM majors succeed in their college-level math and statistics studies. The metric for success in math classes is typically a grade of "B" or higher. We want to understand to what degree the results and outcomes of our work depend on our teaching practices and materials, and to what degree student success is due to differences in background, including prior knowledge, skills, preparation, abilities, etc. To this end, we set out to research commonly accepted college entry exams. Because we have access to educational systems in two counties — the United States (US) and the Russian Federation (RF) — we have compared cross-country approaches to identifying the educational readiness of university STEM majors studying math.

We analyzed the existing research on test results such as the SAT and ACT in the United States and the Unified Country Exam (UCE/ЕГЭ) in the Russian Federation. To our surprise, the goals of each country exams differ significantly. The SAT, for example, is used to measure literacy, numeracy and writing skills needed for academic success in college. Likewise, the ACT is used to measure high school students' general educational development as well as ability to complete college-level work. The UCE, by contrast, assesses knowledge acquired in school, but considers as its main goal providing students with equal opportunity and exam transparency. It is also intended to minimize the possibility of college entry-related corruption.

Several research studies show a good correlation between SAT/ACT grades and success in college studies (Willingham, 1990, Young, 2001, Hezlett et al, 2001, Radunzel, Noble, 2012, Allen 2005).

The predictive power of UCE results for success at college studies is statistically and significantly less (SAT~70+% vs UCE~38%). To make sense of the difference in results, we investigated the exam problems and exam structure in question. The SAT and ACT usually contain a larger number of simpler problems, while the UCE contains a smaller number of problems divided into two parts: A basic section devoted to very simple problems, and an advanced section devoted to some very challenging problems. The latter section is intended for students seeking to enter a STEM program. To some extent, it is similar to SAT subject tests. It differs, however, in the complexity of problems contained therein.

The UCE is meant to create equal opportunity for gifted students of varying educational backgrounds across RF regions. Being gifted does not guarantee success in college math studies. We conclude that college entry exams in different countries (US and RF) are aligned with their respective goals.

In the US, exams concentrate on identifying success in college studies. In the RF, exams concentrate on identifying gifted students.

The SAT/ACT and UCE exams, while similar in form, are based on quite different concepts and, hence, provide quite different outcomes. We conclude that in order to estimate the quality of STEM education we need to take into account the predictive power of college entrance exams.

The paper proceeds as follows. In the introduction section we outline the problem and provide a brief overview of results and a conclusion. Section 2 provides necessary background with a brief historical overview of the SAT, ACT, and UCE exams; gives necessary references to the papers analyzing the tests' college success predictive power (CSPP) and summarizes CSPP per exam. Section 3 describes the structure of the exams and provides typical questions of varying relative complexity per exam and focuses on problem complexity and quantity and time allocated per exam. Section 4 cites statistical analysis of the CSPP, analyses problem complexity and quantity as reasons for possible outcome and offers a hypothesis as to why such a difference in the predictive power exists. Section 5 summarizes our paper results and conclusions.

2. CURRENT COLLEGE ENTRY EXAMS IN USA AND RUSSIAN FEDERATION

We study three main entrance exams available to us from two countries, USA and Russian Federation (RF) as well as their prediction capabilities of future college success.

2.1 USA Entrance Exams: SAT, ACT

In the USA there are two main college entrance exams - the SAT and ACT. We briefly describe them here.

2.1.1 SAT

As defined in (Kobrin, Michel, 2006), "The College Board is a not-for-profit membership association whose mission is to connect students to college success and opportunity. As defined in (Kobrin, Michel, 2006, Goldman, 1976), "The College Board is a not-for-profit membership association whose mission is to connect students to college success and opportunity. Founded in 1900, the association is composed of more than 5,000 schools, colleges, universities, and other educational organizations. The SAT Reasoning Test™ (SAT®) is the most widely used standardized test for college admissions.". The first Scholastic Aptitude Test (SAT) was administered on June 23, 1926 to 8,040 students. The SAT was designed primarily to assess aptitude for learning rather than mastery of subjects already learned. In 2019 over 2 million senior students took the SAT.

The SAT also provides separate subject tests including: foreign languages (Spanish, French, Chinese, Italian, German, Modern Hebrew, Latin, Japanese, Korean), English, physics, chemistry, biology, mathematics (I and II) and history (US and World).

2.1.2 ACT

Another popular college entrance exam in the US, introduced in 1959, is the American College Test (ACT). According to the Princeton Review (*What is ACT?,* 2020, Goldman, 1976) "the purpose of the ACT test is to measure a high school student's readiness for college and provide colleges with one common data point that can be used to compare all applicants." Additionally, the new test should be used not just for admissions but placement as well. It should primarily be useful as an indicator of academic preparation, i.e., it should be an *achievement* test (Sawyer, 2010). Over 75,000 senior students took the ACT in 1959. In 2019 almost 1.8 million seniors took the ACT.

In short, the ACT is more an entrance exam than a "college readiness" predictor. Like the high school GPA (HSGPA), it is more a measure of student's current knowledge base.

2.2 RF Entrance Exam: Unified Country Exam

Introduced in 2009, the Unified Country Exam (UCE/ЕГЭ) in RF is both part of the final certification for educational programs in secondary general education and an entry exam to the higher education institutions. The exam has two mandatory parts: Language (Russian) and Mathematics as well as voluntary parts, which might be needed for entering various Universities (История ЕГЭ, 2020). It still causes a lot of controversy and debate among scientists, university professors, teachers, parents etc. (Неретин, 2016). It has been adopted, nevertheless, by the state and currently the results of the exam are one of the main criteria for enrolling students in universities.

Two of the UCE's mandatory sections - Language (Russian) and Mathematics - are now needed to get one's high school diploma and to apply to a university. Most universities also require some voluntary parts, which include foreign languages (English, German, French, Spanish, Chinese), physics, chemistry, biology, geography, literature, history, basics of social sciences and computer science. The closest US analogy to these voluntary parts are the SAT Subject Tests.

3. STRUCTURE AND CONTEXT COMPARISON OF THE COLLEGE ENTRY EXAMS

In this section we compare structural organization and complexity of the test problems. We will concentrate more closely on the Math part of the exams (Sánchez, 2020).

3.1 SAT

The SAT (excluding SAT Subjects) has three required sections and one optional, with each required section grade in interval 200-800 points:

- Reading, 52 questions, 65 mins.
- Writing, 44 questions, 35 mins.
- Math (includes Science and History/Social Sciences sub-scores), 58 questions, 80 mins.
- Essay (optional), 1 question, 50 mins.

The required SAT test math section contains 58 questions of approximately three levels of complexity. The SAT's typical simplest problems (Stiggins, 1989):

1.   If (x-1)/3=k and k=3, what is the value of x?
2.   If y=kx, where k is a constant, and y=24 when x=6, what is the value of y when x=5?

The SAT's typical mid-range problems:

1.   3x+4y=−23
      2y−x=−19

What is the solution (x,y) to the system of equations above?

A food truck sells salads for $6.50 each and drinks for $2.00 each. The food truck's revenue from selling a total of 209 salads and drinks in one day was $836.50. How many salads were sold that day?

The SAT's typical difficult problems:

1.   If (ax+2)(bx+7)=15x2+cx+14 for all values of x, and a+b=8, what are the two possible values for c?
2.   For a polynomial p(x), the value of p(3) is −2. Which of the following must be true about px()?

A.   x−5 is a factor of p(x).
B.   x−2 is a factor of p(x).
C.   x+2 is a factor of p(x).
D.   The remainder of p(x) divided by x−3 is −2.

3.2 ACT

The ACT has four required sections and one optional of the following structure and time allocation (Saupe 2000):

- Language (English), 75 questions, 45 min.
- Math, 60 questions, 60 mins.
- Reading, 40 questions, 35 mins.
- Science, 40 questions, 35 mins.
- Writing (optional)

All the math questions are five-choice, multiple-choice questions. Total ACT score runs up from 1 to 36.

3.3 UCE

The UCE has two required sections and many optional sections (much like the SAT Subject Tests). They are administered one subject per day:
- Language (Russian), 39 questions, 210 mins.
- Math; basic level: 20 questions, simplest problems, 180 mins; profile level: 19 questions, 8 multiple choice, simplest problems, 4 short answer average problems, 7 difficult problems extended answer, total 235 mins.

The UCE typical simplest problems:

1.   Calculate: 2.4(6.7-3.2)
2.   Ivan earned 20,000 rubles last month. He paid 13% of federal tax. How much money does he have left after paying the tax?

The UCE typical mid-range problems:

1.   If $S=v_0t+at/2$ what is the value of S when $v_0=6$, $t=2$, $a=-2$?
2.   Find the negative solution of $x^2-x-6=0$

The UCE typical difficult problems:
1.   In a cylindrical glass, water reaches a height of 80 cm. The water is poured into another cylindrical glass, whose radius is four times more than the first one. What height, in centimeters, will the water reach in the new glass?
2.   The rectangle is divided into four smaller ones by two straight linear cuts. The perimeters of three of the smaller rectangles, starting with the upper left and counting clockwise are equal to 24, 28 and 16 respectively. Find the perimeter of the fourth rectangle.

Typical UCE problems appear to be harder than problems from the SAT and ACT. More time is allocated to their solution. The SAT and ACT tests basically cover the same knowledge/skills base. However, the SAT has much finer granularity with scores ranging 400-1600 points, while the ACT scores range 1-36.

The UCE test covers a knowledge/skills base in mathematics similar to the SAT and ACT (*Демоверсии ЕГЭ 2020 по математике*, 2020).

3.4 Tests Widespread

Both SAT and ACT tests are almost equally popular with about 2 million US students who took the SAT during high school in 2018 compared with about 1.9 million who took the ACT during high school in 2018.

The UCE test was administered to about 0.75 million RF students in 2018 (see also Clinedinst, 2011).

4. COMPARISON OF THE UNIVERSITY SUCCESS PREDICTIVE POWER OF SAT VS UCE

In this section we discuss CSPP across all tests.

4.1 Methodology

We match the test result with the first-year college GPA (FGPA) as the criterion of the test's predictive power. When comparing the predictive power of the tests, we observed that different tests tend to provide different measures for describing CSPP. A little work is done on comparison of the CSPP across different exams (Noble, 2007).

Typically, the R^2 coefficient from linear regression both for single and multiple variables (for that methodology see (Koretz, et al, 2016, Miles, Shevlin, 2001, Noble, 2002) is used. In (Cohen, 1988) a rule of thumb was suggested for evaluation of correlation. If absolute value of correlation exceeds .50 it is large, if absolute value is between .30 and .50 it is medium, and if absolute values less than .30 it is small. In addition to general issues using R^2 as a measure of prediction power (Miles, Shevlin, 2001), referred here papers used different units and scales. Thus, just normalization of results does not make the outcomes comparable (Green, 2012).

We use the match of the test result and FGPA approach, and re-evaluated data provided in the cited papers to compute the match measure of CSPP (Tinto, 2002), and, hence, compare similar results (Lotkowski 2012, Moore, 2009).

4.2 SAT

Since 1950s, multiple research studies have been conducted that looked at the SAT score as a predictor of student college performance. The earlier studies are summarized in (Fishman, Pasanella, 1960, Willingham, 1990). One recent study was performed by (Young, 2001). Another work worth mentioning is (Hezlett et al, 2001), which presented a comprehensive meta-analysis of about 3,000 validity studies covering more than one million students. The Hazlett study concludes that the SAT is a reasonable predictor of FGPA ranging from .66 to .85. It is summarized in the following table (Kobrin, Michel, 2006):

Table 1. Accuracy Rates for Logistic Regression Models for Total Sample

Success Criterion Level FGPA	Model	Accuracy Rate Overall
2.0	HSGPA Only	87.0
	SAT Only	87.1
	SAT&HSGPA	87.0
2.5	HSGPA Only	72.4
	SAT Only	71.4
	SAT&HSGPA	73.7
3.0	HSGPA Only	67.5
	SAT Only	66.2
	SAT&HSGPA	69.7
3.25	HSGPA Only	73.8

	SAT Only	73.8
	SAT&HSGPA	75.7
3.5	HSGPA Only	83.3
	SAT Only	83.6
	SAT&HSGPA	84.0
3.75	HSGPA Only	99.3
	SAT Only	99.3
	SAT&HSGPA	99.3

The SAT is mostly a better predictor of college success than an average HSGPA. More detailed results stratified by various socio-economical, racial and gender groups can be found in (Kobrin, Michel, 2006).

4.3 ACT

A good source for data on ACT as a college success predictor is (Radunzel, Noble, 2012, Allen 2005). In particular, it gives the following data for ACT/HSGPA predictive power, comparable to the above data for SAT. The ACT is a reasonable predictor of FGPA ranging from .58 to .68 (see also Schmitt, 2009).

Table 2. ACT Predicting Levels at Four-Year Institutions

Success Criterion Level FGPA	Model	Accuracy Rate Overall
3.0	HSGPA Only	70
	ACT Only	68
	ACT&HSGPA	71
3.25	HSGPA Only	63
	ACT Only	64
	ACT&HSGPA	67
3.5	HSGPA Only	56
	ACT Only	61
	ACT&HSGPA	63
3.75	HSGPA Only	50
	ACT Only	58
	ACT&HSGPA	59

4.4 UCE

According to (*Психолого-педагогический анализ успеваемости студентов по высшей математике,* 2018, Хавенсон, Соловьева, 2014, Польдин, 2011), the UCE is not as good a predictor of FGPA as the SAT or ACT. The sample sizes used in (Польдин, 2011) are not large enough though to make reasonably definite conclusions compared with the above data for the SAT and ACT. The UCE is a predictor of FGPA with predictive power ranging from .33 to .48 (see also Sackett, 2009).

Table 3. UCE Predicting Levels at Four-Year Institutions

Success Criterion Level FGPA	Model	Accuracy Rate Overall
3.0	UCE Only	35
2.0	UCE Only	33
1.0	UCE Only	48

5. CONCLUSION

In order to understand to what degree the results and outcomes of our work depend on our teaching practices and materials, and to what degree student success is due to differences in background, including prior knowledge, skills, preparation, abilities, etc., we considered the commonly accepted college entry exams.

We have compared cross-country approaches to identifying the educational readiness of university STEM majors studying math available to us via US and RF college entry exams. We analyzed test structure and content. For the US, tests contain more questions, with time per question under 1 minutes 30 seconds, and with questions being relatively simple. For the RF, tests contain fewer questions with over 10 minutes allocated per question, with questions being relatively more difficult.

In the US the main college entry exams, the SAT and ACT, are used to measure literacy, mathematical and writing skills necessary for academic success in college. In the RF the UCE is used to provide equal opportunity for students as well as to fight corruption.

In the US, exams concentrate on identifying success in college studies. In the RF, exams concentrate on identifying gifted students. According to our analysis, the tests are based on different goals which they respectively satisfy.

REFERENCES

ACT, 2012. *ACT scores and postsecondary degree attainment: What's the connection?* ACT Information Brief No. 2012-11). Iowa City, IA.

Allen, J., & Sconing, J., 2005. *Using ACT Assessment scores to set benchmarks for college readiness*. (ACT Research Report No. 2005-3). Iowa City, IA: ACT, Inc.

Clinedinst, M.E., Hurley, S.F., et al, 2011. *2011 State of College Admission*. Alexandria, VA: National Association for College Admissions Counseling.

Cohen, J., 1988, "*Statistical Power Analysis for the Behavioral Sciences (2nd ed.)*", Hillsdale, NJ., Erlbaum

Fishman, J.A., & Pasanella, A.K., 1960, "*College admission selection studies*", Review of Educational Research, 30(4), 298–310

Goldman, R.D., & Widawski, M.H., 1976. *A within-subjects technique for comparing college grading standards: Implications in the evaluation of college achievement*. Educational and Psychological Measurement, 36, 381–90.

Green, C., & Radwin, D., 2012. *Characteristics of associate's degree attainers and time to associate's degree*. (Web Tables NCES 2012-271). Washington, DC: U.S. Department of Education.

Hezlett, S.A., et al, 2001, "*The effectiveness of the SAT in predicting success early and late in college: A comprehensive meta-analysis.*",The annual meeting of the National Council on Measurement in Education, Seattle, WA

Kobrin, L.L., Michel, R.S., 2006, "*The SAT® As a Predictor of Different Levels of College Performance*", https://files.eric.ed.gov/fulltext/ED563073.pdf

Koretz, D., Yu, et al, 2016, "*Predicting Freshman Grade Point Average from College Admissions Test Scores and State High School Test Scores.*", AERA Open

Lotkowski, V.A., Robbins, et al, 2004. *The role of academic and non-academic factors in improving college retention*. Iowa City, IA: ACT, Inc.

Miles, J., Shevlin, M., 2001, "*Applying regression and correlation: A guide for researchers.*", London, SAGE.

Moore, C., Shulock, N., et al, (2009. *Steps to success: analyzing milestone achievement to improve community college student outcomes*. Sacramento, CA: Institute for Higher Education Leadership & Policy. National Governors Association Center for Best Practices & Council of Chief State School

Noble, J., & Radunzel, J., 2007. *College readiness = college success beyond the first year*. Paper presented at the Annual Forum of the Association for Institutional Research, June 2–6, Kansas City, Missouri. 51

Noble, J., & Sawyer, R., 2002. *Predicting different levels of academic success in college using high school GPA and ACT composite score*. ACT Research Report Series, Iowa City, IA: ACT, Inc.

Radunzel, J. Noble, J., 2012, "*Predicting Long-Term College Success through Degree Completion Using ACT® Composite Score, ACT Benchmarks, and High School Grade Point Average*", http://www.act.org/content/dam/act/unsecured/documents/ACT_RR2012-5.pdf

Robbins, S., Allen, J., et al, 2006. *Unraveling the differential effects of motivational and skills, social, and self-management measures traditional predictors of college outcomes*. Journal of Educational Psychology, 98, 598-616.

Sackett, P.R., Kuncel, N.R., et al 2009. *Does socioeconomic status explain the relationship between admission tests and postsecondary academic performance?* Psychological Bulletin, 135(1), 1-22.

Sánchez H., Comeaux E., et al, 2020. *Report of the UC Academic Council Standardized Testing Task Force*, https://senate.universityofcalifornia.edu/_files/underreview/sttf-report.pdf

Saupe, J.L., & Curs, B.R., 2008. *Deriving enrollment management scores from ACT data*. (IR Applications No. 16). Tallahassee, FL: The Association for Institutional Research.

Sawyer, R., 2010. *Usefulness of high school average and ACT scores in making college admission decisions*. (ACT Research Report No. 2010-2). Iowa City, IA: ACT, Inc.

Schmitt, N., Keeney, J, et al, 2009. *Prediction of 4-year college student performance using cognitive and noncognitive predictors and the impact on demographic status of admitted students*. Journal of Applied Psychology, 94(6), 1479-1497.

Stiggins, R.J., Frisbie, D.A., et al, 1989. *Inside high school grading practices: Building a research agenda*. Educational Measurement: Issues and Practice, 8(2), 5–14.

Tinto, V., 2002. *Enhancing student persistence: Connecting the dots*. Paper presented at the Optimizing the Nation's Investment: Persistence and Success in Postsecondary Education conference sponsored by the Wisconsin Center for the Advancement of Postsecondary Education.

What is ACT?, 2020, https://www.princetonreview.com/college/act

Willingham, W., Lewis, C., et al, 1990. *Predicting college grades: An analysis of institutional trends over two decades*. Princeton, NJ: Educational Testing Service.

Young, J.W., 2001, "*Differential validity, differential prediction, and college admission testing: A comprehensive review and analysis.*", College Board Research Report N 6. New York: The College Board

Демоверсии ЕГЭ 2020 по математике, 2020, https://4ege.ru/matematika/58234-demoversii-ege-2020-po-matematike.html

История ЕГЭ, 2020, http://edu.glavsprav.ru/spb/ege/history/

Неретин Ю.А., 2016, "*ЕГЭ и агония математики в школе*", Математическое образование. № 4 (80). С. 2-14, http://mi.mathnet.ru/mo583

Польдин, О.В., 2011, "*Прогнозирование успеваемости в ВУЗе по результатам ЕГЭ*", Прикладная Эконометрика, #1(21), http://www.appliedeconometrics.ru/general/upload/articles/pe2111-56-renamed.pdf

Психолого-педагогический анализ успеваемости студентов по высшей математике, 2018, http://adlog.narod.ru/st/statia13.htm

Хавенсон, Т.Е., Соловьева А.А., 2014, "*Связь результатов Единого государственного экзамена и успеваемости в вузе*", Вопросы образования, https://vo.hse.ru/2014--1/117875470.html