Journal of the National College Testing Association
2015/Volume 1/Issue 1

Evaluating a Retest Policy at a North American University

CINDY L. JAMES, PhD
Thompson Rivers University

Retesting for admission or placement in higher education is quite common; hence, it is important to ensure that retest policies endorsed by post-secondary institutions are judicious. This is especially true since policies do vary with some institutions allowing no retests, others allowing unlimited retests, and yet others having preconditions such as limiting the number of retests and/or time between tests or requiring an intervention before a retest. Utilizing assessment data, this study evaluated the efficacy of one North American university's retest policy of allowing one retest with no preconditions. The findings revealed that scores on five of the ACCUPLACER tests (three English and two math) increased for the majority of the potential students who retested. Moreover, on average, retest scores were significantly higher than original test scores, with the effect size for all of these differences being large. This study also revealed that—for the English tests—lower scores on the original tests were associated with larger score gains on the retests and vice versa. Based on these findings, it seems that the retest policy at this particular institution is efficacious in that it facilitates an acceptable practice.

INTRODUCTION

Due to the high stakes nature of standardized testing in higher education, it is quite common to have candidates retake admission or placement tests in order to improve their test scores. To enhance candidates' probability of increasing their test scores and to prevent "*serial*" testing, educational institutions commonly employ a retest policy. Usually such policies establish preconditions for retesting such as limiting the number of retests or the time between tests or requiring some type of intervention before allowing retests. In many cases, especially with high-stakes exams, retest policies are established by the testing agency, but in some cases post-secondary institutions are able to set their own retest policies.

No matter how they are established, such policies do appear to vary. For example, candidates may write the ACT twelve times in total (ACT, 2014). For the Graduate Record Exam (GRE) General test, candidates may write the test up to five times per twelve-month period, but there must be at least twenty-one days between retests (Educational Testing Service, 2014). Similarly, the Graduate Management Admission Test (GMAT) may be written up to five times in any twelve-month period but only once per thirty-one–day period (Graduate Management Admissions Council, 2014). Canadidates for the Law School Admissions Test (LSAT) may not take the LSAT more than three times in any two-year period (Law School Admission Council, 2014a), and for the Medical College Admissions Test (MCAT), candidates may

Cindy L. James is the senior coordinator of the Assessment Centre at Thomson Rivers University, Kamloops, BC, Canada. Email: cjames@tru.ca.
Phone: 250-828-5471

take the exam a maximum of three times during a single testing year, four times during a two consecutive-year period and seven times in total (Association of American Medical Colleges, 2015).

In comparison, for tests such as the SAT and ACCUPLACER, there does not appear to be any limit on retests. However, for ACCUPLACER the College Board does encourage institutions to establish their own retest policy. Based on a survey conducted by the College Board, institutional retest policies for ACCUPLACER appear to vary substantially (College Board, 2014a). Specifically, some institutions do not allow any retests, some limit it to one retest per semester, year or two-year period, and others allow unlimited retesting.

With such diversity, it is difficult to determine, which, if any, of these policies is most effective. However, by utilizing assessment data, educational institutions could examine such policies by measuring score differences and other factors related to retesting, such as the frequency of retesting. For example, Andrews and Ziomek (1998) reported that approximately 36% of the ACT-tested graduating class of 1993 wrote the ACT more than once during their junior and senior academic years. In most cases, there was an increase in the mean score gain from the first test to the retest, but the size of the gain depended on the score obtained in the first testing. Hence, the largest gains on the second testing were made by examinees with the lowest scores on first testing, and the smallest gains were made by examinees with higher composite scores on the first testing (Andres & Ziomek, 1998).

According to GMAC, approximately 20% of candidates retake the GMAT with an average score gain of thirty-three points on a 200–800 scale (Rudner, 2012). However, similar to the ACT, the gains do differ by score group: those who scored higher on the first test gained, on average, less on the retest than those who scored lower on the first test (Rudner, 2012). Similar rates of retesting were reported for the LSAT, with 26.1% of candidates rewriting once and 7.1% rewriting more than twice in 2012–2013 (LSAC, 2014b). As with the GMAT and ACT, the size of the score gains varied somewhat by first test scores, but overall, increases in scores were reported for the majority of students who retested. For example, 67% of the 1,256 candidates who scored 150 on a previous LSAT exam improved their scores on a retest in 2012–2013 (LSAC, 2014b).

Based on information provided by the College Board, SAT scores improved for 55% of students who wrote the SAT as juniors and again as seniors (College Board, 2014b). On average, the combined scores on reading, mathematics, and writing increased by approximately forty points (College Board, 2014b). A study by Bostian and Henry (2014) also reported increases on rewrites for ACCUPLACER. Specifically, on retests of Reading Comprehension, average gains ranged from 6.6 to 13.0, and for Sentence Skills, average gains ranged from 6.6 to 13.6 (both out of 120). The average gains on retests for both math tests (also out of 120) were even higher, ranging from 9.5 to 18.6 for Elementary Algebra and from 20.3 to 25.4 on Arithmetic (Bostian & Henry, 2014). The average time between testing sessions differed somewhat by test, but ranged from 171 days to 275 days (Bostian & Henry, 2014).

Although these studies provide some valuable insight into retesting, further investigation, especially at the institutional level, is needed. This study was designed to do exactly that by exploring ACCUPLACER retesting activity at one North American university to determine whether the retesting policy of allowing candidates to retake ACCUPLACER once without any preconditions was judicious. To evaluate this policy, assessment data were examined to determine the frequency of retests and the score differences between first tests and retests. The frequency data were collected to serve as an indicator of the impact of the retesting policy. The changes in test scores—direction and magnitude—were calculated to validate the assumption that retests provide candidates with an opportunity to improve their test scores. The latter calculations also were used to assess the practice of allowing retesting without any preconditions such as limiting the time between tests or requiring some type of intervention before allowing a retest.

METHODS

This study was conducted at a comprehensive, public university that educates approximately 13,000 learners on campus and another 11,500 by distance and online (TRU, 2015). On-campus students can choose from 140 programs offering certificates, to bachelor and graduate degrees in trades, traditional academics, and professional career options (TRU, 2015).

Entry Assessment Tool—ACCUPLACER

ACCUPLACER is a battery of tests that assesses students' academic skills in English, mathematics, and computer skills. Approximately 1,500 secondary and post-secondary institutions utilize ACCUPLACER test scores for advising, placement, and/or admission purposes (College Board,

2014c). Currently, these institutions administer 8 million tests to more than 2.5 million students annually (College Board, 2014c).

ACCUPLACER, as a web-based system, offers reduced testing time, enhanced security features, immediate feedback, and flexible testing sessions. It also is a *computer-adaptive* testing system, whereby the tests automatically adjust to the skills of the individual examinee. Consequently, it is capable of assessing a wide range of student abilities.

The primary tests used at this university include Reading Comprehension, Sentence Skills, Arithmetic, Elementary Algebra, and WritePlacer. A detailed description of the content of each test is available on the College Board website (https://accuplacer.collegeboard.org/students/accuplacer-tests). The scores for Reading Comprehension, Sentence Skills, Arithmetic, and Elementary Algebra range from 0 to 120, and scores for the WritePlacer range from 0 to 8.

For the multiple choice tests, the bank of questions is significantly large. Specifically, there are 537 questions for Reading Comprehension, 297 questions for Sentence Skills, 424 questions for Arithmetic, and 517 questions for Elementary Algebra (K. Montognese, personal communication, January 20, 2015). Moreover, these items are refreshed on a regular basis, such that after any question reaches a certain level of exposure, it is retired and replaced by a new one (K. Montognese, personal communication, January 21, 2015). Hence, for retests, the likelihood of being exposed to similar questions for the multiple choice tests is quite low. As for the WritePlacer test, thirteen prompts are available. At this institution, one of these prompts is randomly assigned during each testing session; thus, the probability of being assigned the same prompt during a retest is also fairly low.

ACCUPLACER Testing Procedure

At this institution, ACCUPLACER is utilized for both placement and admissions testing. Applicants applying for certain trades (e.g., automotive, welding), technology (e.g., electrical, electronics), and career programs (e.g., administrative assistant, health care assistant) are required to complete the testing as a part of the admissions process. Applicants requiring upgrading prior to entering other university programs may also take the ACCUPLACER test, but such testing is not mandatory. For each program, the number and type of ACCUPLACER tests required do vary. Some programs require only Reading Comprehension and Arithmetic, and others require both of these tests plus Sentence Skills and Elementary Algebra (see Appendix I for a list of programs, tests, and cutscores required). All the tests are untimed except for WritePlacer, which has a one-hour time limit.

The tests are administered on a weekly basis at the central campus and on demand at proctored sites off campus. After testing has finished, applicants are given a copy of their results and informed about the retest policy. This policy states, "An applicant may rewrite one or more of the ACCUPLACER tests after their initial attempt. It is highly recommended that the applicant spend several weeks reviewing the subject matter before attempting the rewrite." If and when applicants retest, they are required to rewrite only the tests in which they did not achieve the necessary cutscore. On average, applicants waited 36 days (*SD = 42.5*) before attempting a retest, but this did vary. For example, one applicant rewrote within one day of his/her first test, and another waited 287 days before rewriting.

Subjects

For this study, data from a five-year period (2010 to 2014 inclusive) were collected. During this time, 5,766 potential students completed the ACCUPLACER tests. The average age of these candidates was 25.5 (*SD = 9.78*) with 62.4% being male, 37.4% female, and .2% unspecified. A total of 878 (15%) students retook an ACCUPLACER test within one year of their first test. The average age of the candidates who retested was 22.5 *(SD = 8.34)* with 69.7% being male and 30.3% being female.

Data Analysis

Since there are no national norms for ACCUPLACER, the descriptive statistics from all the candidates completing one or more tests from 2010 through 2014 were tabulated to provide institutional norms. Further analysis involving scores from retesting activity also were tabulated. To assess the distribution of ACCUPLACER test and retest scores, the kurtosis, skewness, and standard errors were calculated for each test on the two different testing occasions. Based on these statistics, the first test and retest scores appeared to follow a normal distribution except for Arithmetic and Elementary Algebra. Consequently, comparisons between first test and retest ACCUPLACER scores were conducted utilizing paired t-tests for the English tests and Wilcoxon Signed Rank tests for the math tests, along with the appropriate calculations for effect sizes. No assumption was made about the direction of the retest scores (increasing or decreasing); therefore, two-tailed tests were conducted with the level of significance set to .05. Correlation analysis between first test scores and changes in test scores were also conducted to determine if lower scores on the initial test related to higher scores on the subsequent test and vice versa. Pearson's *r* was calculated for the English tests and

Spearman's *rho* was calculated for the math tests, with the level of significance set at .05.

RESULTS

The institutional norms for ACCUPLACER testing between 2010 and 2014 are provided in Table 1. In total, 5,766 candidates were tested during this time, but as noted in the previous section, the combination of tests completed by each candidate varied depending on the program they were applying to. Since most programs required Reading Comprehension, it was the most common test taken, followed by the Arithmetic, Elementary Algebra, Sentence Skills and WritePlacer (Table 1).

TABLE 1
Institutional Norms: Descriptive Statistics of ACCUPLACER Testing from 2010 to 2014

ACCUPLACER Test	*N*	% of Applicants Taking Each Test	Mean	*SD*	Minimum	Maximum
Reading Comprehension	5,172	89.7%	80.05	20.94	27	120
Sentence Skills	3,193	55.4%	83.13	21.37	29	120
WritePlacer	1,621	28.1%	4.31	1.54	0	8
Arithmetic	4,693	81.4%	75.03	30.83	20	120
Elementary Algebra	3,514	60.9%	57.86	28.20	21	120

N = 5,766

Of the 5,766 potential students who were tested, 878 (15%) completed a retest within one year. As mentioned earlier, these students were required to retake only tests in which they did not achieve the necessary cutscores. Consequently, in most cases, applicants retook only one test (66%); however, a fair number took two tests (24%), and some attempted three (8%) or four (2%) tests. Out of 3,514 candidates who completed the Elementary Algebra test initially, 345 retested (9.8%) making it the most frequently repeated test. Arithmetic was second with 358 out of 4,693 (7.6%) candidates retesting, and Reading Comprehension was close behind with 377 out of 5,172 (7.3%) retesting. Tests with the lowest frequencies of retests were WritePlacer with 74 of the 1,621 (4.6%) candidates retesting and Sentence Skills with 121 out of 3,193 (3.8%) candidates retesting.

The descriptive and comparative statistics of the test scores for the applicants who retested are provided in Table 2.

TABLE 2
Comparison of Test and Retest Scores: Paired t-test and Wilcoxon Signed Ranks Tests Results

		FIRST TEST				RETEST				DIFFERENCE		
ACCUPLACER Test	*n*	Mean	*SD*	Skewness	Kurtosis	Mean	*SD*	Skewness	Kurtosis	Mean	Statistic	Effect Size
Reading Comprehension	377	55.71	12.59	.24	–.14	67.99	16.34	.01	–.14	12.28	t =16.93*	.87
Sentence Skills	121	59.34	12.43	.40	.63	73.26	16.75	.18	–.40	13.92	t = 9.73*	.88
WritePlacer	74	3.86	1.23	–.32	–.84	4.85	1.19	.295	.62	.97	t = 8.06*	.94
Arithmetic	358	46.34	13.97	.92	1.99	68.46	24.74	–.069	–1.06	22.12	Z = –13.56*	.72
Elementary Algebra	345	30.34	8.94	2.92	6.85	44.29	18.28	1.035	.54	13.95	Z = –13.29*	.72

*p <.001, two-tailed.

Based on the paired t-tests, the means from the retests were significantly higher than those from the first tests for all three English tests. The effect sizes for these differences were large (Cohen, 1988). Similarly, based on the Wilcoxon Signed Rank tests, there were significant differences in test and retest scores for the two math tests, with both exhibiting large effect sizes.

Table 3 shows the breakdown of test scores, as a function of whether they increased, decreased, or did not change upon retesting.

TABLE 3
The Distribution of the Changes in Retest Scores

		Applicants					
ACCUPLACER Test	***n***	**Scores Increased**		**No Change**		**Scores Decreased**	
Reading Comprehension	377	307	81%	6	2%	64	17%
Sentence Skills	121	95	78%	2	2%	24	20%
WritePlacer	74	50	67%	19	26%	5	7%
Arithmetic	358	290	81%	7	2%	61	17%
Elementary Algebra	345	280	81%	19	6%	46	13%

With the exception of WritePlacer, on which only two out of every three students saw their scores improve, approximately 80% of students saw their ACCUPLACER scores improve upon retesting. Based on the individual data, the majority of the score gains were large (i.e., greater than 10 points on the multiple choice tests), whereas the majority of the score decreases were relatively small (i.e., less than 10 points on the multiple choice tests).

Based on the correlation analysis, there were significant negative associations between initial test scores and the differences in tests scores for all three English tests, but not for the math tests (Table 4). The significant, negative correlations indicate that lower scores on initial tests related to larger gains on the retests, and higher initial test scores related to smaller gains on the retests for Reading Comprehension, Sentence Skills, and WritePlacer.

TABLE 4
Correlation Analysis of First Test Scores and Score Differences for the Five ACCUPLACER Tests

ACCUPLACER Test	***n***	**Correlation Coefficient**	***R*-squared**	***F***	**Sig (2-tailed)**
Reading Comprehension	377	$r = -.25$	.064	25.77	.000
Sentence Skills	121	$r = -.31$	.096	12.71	.001
WritePlacer	74	$r = -.47$	.207	19.998	.000
Arithmetic	358	$r_s = -.064$			.077
Elementary Algebra	345	$r_s = -.027$			.612

DISCUSSION

In higher education, retest policies for admission and placement testing often vary. Many of these policies are established by the testing agency, but others are established by educational testing centers operating within the post-secondary system. For the latter, evaluating the efficacy of any such policy is crucial. Generally, the assumption behind retesting is that it allows candidates the opportunity to improve their test scores, but do scores actually improve and, if so, by how much? To explore this issue, this study utilized assessment data to evaluate a North American University's

retest policy of allowing potential students to rewrite ACCUPLACER tests once without any preconditions.

Based on the results at this particular institution, retesting is not common practice, as only 15% of candidates rewrote one or more of the ACCUPLACER tests within one year of completing their first test. Overall, the test scores increased for the majority of the potential students who retested. However, test scores remained the same or decreased for 19% of the candidates who rewrote Reading Comprehension, Arithmetic, and Elementary Algebra, and for 22% of the candidates who rewrote Sentence Skills. As for the WritePlacer test, 33% of the candidates did not improve their scores on the retests. From the data collected, it is not possible to explain the changes, or lack thereof, in test scores. However, for the WritePlacer it is worth noting that the scoring scale is much smaller than for the other tests, an eight-point scale for WritePlacer as compared to a 120-point scale for the multiple choice tests. With fewer points, performance differences on WritePlacer have to be much greater to register on the score scale. This likely explains why a higher proportion of the retest scores for WritePlacer tended to remain the same in comparison to the other ACCUPLACER tests.

Even though not all individual scores improved, the mean scores on all five tests improved significantly for those who retested. Arithmetic scores exhibited the greatest average gain of twenty-two points, while score gains on the other three multiple choice tests—Reading Comprehension, Sentence Skills and Elementary Algebra—ranged from twelve to fourteen points. Changes in the WritePlacer test scores were also significant, increasing, on average, by one point on an eight-point scale. These results are somewhat similar to those reported by Bostian and Henry (2014), especially in terms of the gains made on Arithmetic. The fact that the score gains varied somewhat by subject matter is not too surprising. Specifically, the substantial improvement of scores on Arithmetic is logical because—in the short term at least—it is probably more effective to review and refresh basic math concepts than it is to do so with other skills, such as reading comprehension and writing. Although this finding may not influence retest policy per se, it may be worth sharing with candidates contemplating retesting.

Similar to the studies by Andres and Ziomek (1998) and Rudner (2012), the magnitude of the gains in this study appear to depend somewhat on the initial scores of the ACCUPLACER tests. Specifically, based on the correlation analysis, lower scores on the first tests were associated with greater gains on the retests, while higher scores on the first tests equated to lesser gains on the retests for Reading Comprehension, Sentence Skills, and WritePlacer. This makes sense, as there would be a ceiling effect for candidates who scored higher on the first test. That is to say, there is not as much room to improve scores if the initial scores are quite high. In comparison, there is greater potential to score higher on subsequent tests when the initial test scores are lower. Again, this finding may not necessarily impact retest policy, but it could be used to advise candidates considering retesting.

Given that the retest scores increased for the majority of the candidates and that, on average, these increases were significant, it would seem that the existing retest policy at this institution is judicious in that it facilitates an acceptable practice. Specifically, it appears reasonable to allow applicants to retest once without any preconditions, such as limiting the time between tests or requiring some type of intervention. It is important to acknowledge that this conclusion is based on the assumption that the applicants' retest scores are valid. This assumption is inherent to the practice of retesting, as it would be incongruous for testing agencies or educational institutions to allow retesting if they did not believe retest scores provided an accurate measurement of candidates' true abilities. Nonetheless, the validity of retest scores is a key issue that merits scrutiny in subsequent studies.

Limitations and Future Research

Although this research provided insights into ACCUPLACER retesting, there are several limitations that hinder generalizations and implications. To start with, as with all local studies, the data collected may not be applicable to other institutions. Hence, similar studies need to be conducted at other post-secondary institutions, especially those with different populations, testing tools, and/or retest policies.

Examining the validity of the retest scores, as already noted, is also necessary. This is important as score improvements—such as the ones revealed in this study—may be a result of test familiarity or measurement error rather than skill development. This could be done by exploring the repercussions of retesting on academic performance. For instance, by comparing the course or program performance of students who retested to those who did not, it would be possible to assess the validity of retest scores, as well as initial test scores. Unfortunately, this type of data was not readily accessible for this study.

At this institution, candidates retaking the ACCUPLACER are not asked about their preparation activities; thus, it is not possible to explore connections between such activities and score variances. Since this type of analysis has the potential to identify the least and/or most effective test preparation strategies, it would be highly recommended that subsequent studies survey applicants about their preparation activities for retests.

Finally, exploring candidates' reactions to retests, and perhaps more importantly, to retest policies, would be worthwhile. For instance, if applicants are not allowed to

retest, does this hinder their performance by increasing stress levels, or does it enhance their performance by motivating them to be better prepared? Although candidates' reactions should not necessarily determine policy, such input certainly could inform policy development or revision.

Conclusion

Utilizing five years of testing data, this study evaluated a North American university's retest policy of allowing one retest without any preconditions. Assessing the frequency of retests and the score differences between tests and retests made it possible to conclude that the retest policy at this institution was acceptable. Whether the same can be said about retest policies at other educational institutions is uncertain unless such institutions conduct similar studies of their own.

REFERENCES

ACT, Inc. (2014). *Retest restrictions*. Retrieved from http://www.actstudent.org/regist/retestrestrictions.html

Andrews, K.M. & Ziomek, R.L. (1998). *Score gains on retesting with the ACT assessment*. ACT Research Report Series 98-7. Iowa City, IA: ACT.

Association of American Medical College (2015). *How often will I be able to take the MCAT?* Retrieved from https://www.aamc.org/students/applying/mcat/faq/421622/how-often-can-i-take-the-mcat-exam.html

Bostian, B. & Henry, K. (2014, June). *Resources for ACCUPLACER test prep*. Paper presented at the 24th annual ACCUPLACER National Conference, Indianapolis, IN.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

College Board (2014a). *ACCUPLACER program manual.* New York, NY: Author.

College Board (2014b). *Retaking the SAT*. Retrieved from http://professionals.collegeboard.com/testing/sat-reasoning/scores/retake

College Board (2014c). *About ACCUPLACER*. Retrieved from http://accuplacer.collegeboard.org/professionals

Educational Testing Service (2014). *2014–15 GRE® Information and Registration Bulletin*. Retrieved from http://www.ets.org/s/gre/pdf/gre_info_reg_bulletin.pdf

Graduate Management Admission Council (2014). *GMAT Handbook*. Bloomington, MN: Author.

Law School Admission Council (2014a). *Frequently Asked Questions: LSAT*. Retrieved from http://www.lsac.org/jd/help/faqs-lsat

Law School Admission Council (2014b). *LSAT repeater data*. Retrieved from http://www.lsac.org/docs/default-source/data-%28lsac-resources%29-docs/repeaterdata.pdf

Rudner, L. (2012). *Behind the scenes: Repeat testing*. Retrieved from http://www.mba.com/global/the-gmat-blog-hub/the-official-gmat-blog/2012/dec/behind-the-scenes-repeat-testing.aspx

Thompson Rivers University (TRU). (2015). *About us*. Retrieved from http://www.tru.ca/about.html

Appendix I: TRU ACCUPLACER Testing Information

	Minimum Score Required for Each Test				
Program	*Reading Comprehension*	*Sentence Skills*	*Write Placer*	*Arithmetic*	*Elementary Algebra*
Aboriginal Health Career	65	60	4	55	30
Applied Business Technology	75	70	—	—	—
Bachelor of Education in Trades & Technology	—	—	—	75	50
Business Diploma	—	—	—	75	50
Construction Trades: Carpentry/Residential/Joinery	65	60	—	65	40
Construction Trades: Piping	75	—	—	65	40
Electrical & Electronics Industrial Electrician/ Instrumentation Mechanic	75	70	—	75	50
Food Training: Culinary Arts & Retail Meat Processing	55	—	—	55	—
Health Care Assistant	65	60	4	—	—
Mechanical Trades: Automotive Service Technician	75	—	—	65	40
Mechanical Trades: Heavy Duty/ Commercial Transport Welding	55	—	—	55	30
Other Trades: Partsperson, Trowel & Water Treatment	55	—	—	55	—
University Entrance-Level English	85	80	6	—	—
University Preparation—English	varies	varies	varies	—	—
University Preparation—Math				varies	varies