

Cabrillo College

**ACCUPLACER validation
and comparison to MDTP**

Terrence Willett
Research Technician
Office of Institutional Research
April 7, 2000

ACCUPLACER validation and comparison to MDTP

Terrence Willett
Research Technician
Office of Institutional Research
April 7, 2000

Summary

- Over 75% of ACCUPLACER placed students were deemed appropriately placed by responding instructors with the exception of Math 4 (Pre-Calculus) students.
- Proportions of placements into the 4 levels of math with ACCUPLACER appear different than historical proportions suggesting the need to adjust cut scores. Specifically, the Arithmetic/Elementary Algebra boundary appears too high and the Elementary Algebra/Intermediate Algebra and the Intermediate Algebra/Transfer level boundaries appear too low.
- ACCUPLACER scores appear to relate modestly well to MDTP but the results are compromised by experimental errors.
- ACCUPLACER has been fully approved by the State Chancellor's Office negating the need for further local validation.
- ACCUPLACER has been associated with data loss and poor technical support.
- Continuation with MDTP is recommended while continuing the search for a replacement computer adaptive math placement instrument.

Introduction

For Math placement, Cabrillo College has long used the Mathematics Diagnostic Testing Project (MDTP) developed as a joint project of the California State University and the University of California in 1977 and first released 1982. In Fall of 1999, a computer adaptive test (CAT) called ACCUPLACER by the College Board was used in its place.

CAT tests take advantage of new technology in the context of Item Response Theory (IRT) to create potentially a more efficient and individualized assessment (Wainer 1990). Classical test theory arose out of the standardization of assessment tests. Prior to this, assessment of skills was often a set of trials posed by a master to an apprentice. In this setting, testing could be individualized to meet the skill level of the examinee. Increased societal complexity led to the mass administration of assessment exams as seen in the civil service exams in ancient China. In the West, the rise of scientific and industrial worldviews favoring quantification and efficiency inspired the development of standardized tests in place of individualized test. By the beginning of the last century in the U.S., the Army and the College Board had both developed standardized exams that administered a common pool of items to all examinees. Since then, we have witnessed the proliferation of standardized assessments in the Armed Forces, education, psychometrics, and industry. While practical to implement, this type of test suffers from unmet assumptions such as parallelism (obtaining the same score from the same examinee on multiple administrations) and equality of variance of errors of measurement between examinees. In addition, for comparison purposes, examinees are all given the same test items or at least tests shown to be of comparable difficulty. These tests are

generally constructed to be of moderate difficulty and poorly differentiate students of very low or very high ability.

Item Response Theory responds to these difficulties by positing a latent ability, θ , that can be assessed given items of known difficulty, b (Lord 1980; Hambleton and Swaminathan 1985, Wainer 1990). IRT also includes parameters to account for item discrimination, a , and guessing, c . These are combined into a 3-parameter model:

$$P(\theta) = c + \frac{1 - c}{1 + e^{-1.7a(\theta - b)}}$$

$P(\theta)$ is the probability of respondent with ability θ to answer correctly a question of difficulty b ,

a is proportional to the slope, $0.425a(1-c)$, referred to as item discrimination,

b = item difficulty

c = guessing parameter

θ = respondent's ability.

The graph of the above logistic model is called an Item Characteristic Curve. The theory holds that ability discrimination is maximized when $P(\theta)$ remains near the inflection point for a bank of items. This point is at 50% when there is no guessing. To maintain maximum discrimination for each item, the set of questions cannot be the same for all as ability levels vary between examinees. In the CAT format, a correct response solicits a more difficult question and an incorrect response a less challenging one to keep $P(\theta)$ near the curve's inflection point until θ can be estimated within an acceptable error range. This takes far fewer questions than a traditional test and so can reduce the resources needed for testing and reduce fatigue for examinees. In addition, this method is theoretically more accurate as it is not based upon the restrictive assumptions of classical test theory such as parallelism and homogeneity of examinee measurement error.

Because of the potential advantages of increased accuracy and decreased examination time, one CAT test, the ACCUPLACER by the College Board, was compared with a paper and pencil version of MDTP to assess its validity and appropriateness for use at Cabrillo. Originally, the purpose was to provide data to the State Chancellor's Office to validate ACCUPLACER as a locally managed test. However, ACCUPLACER recently received full approval from the Office disarming the need to provide data for validation purposes. Now this report serves as an aid to Math faculty in deciding which test they feel is best suited for placing students.

Methods

To help determine if students were placed appropriately by ACCUPLACER in Fall 1999 using cut scores determined by the Assessment Center, instructors were asked to evaluate how prepared they felt their students were for their class on a 5 point scale ranging from "very strong ability" to "very weak ability" (California Community College Matriculation Resource Manual 1998). 25 instructors evaluated a total of 321 students in Math 154 (Elementary Algebra), Math 152 (Intermediate Algebra), Math 12 (Statistics), Math 13 (Finite Math), Math 18 (Business Calculus), Math 4 (Pre-Calculus). The Resource Manual suggests a threshold of at least 75% of students being placed in the 3

highest levels of preparation to have confidence in the appropriateness of placement by an instrument.

To further determine validity and set proper cut scores, the MDTP Intermediate Algebra test was compared with the ACCUPLACER Algebra test. The Cabrillo College Assessment Center double tested 58 Cabrillo College students who were compensated \$10 for their participation but due to data collection errors, only 52 students provided usable data. Students were not informed which test was the experimental and which was the control instrument. All 52 students took the ACCUPLACER elementary algebra computer adaptive test (CAT). Of those, 38 took only the MDTP Algebra Readiness test, 14 took only the MDTP Intermediate Algebra test, and one took both tests. All the MDTP tests are paper and pencil (PnP) tests. Due to errors in administering the research protocol, students were allowed to choose which version of the MDTP they wanted to take rather than all being given the Intermediate Algebra level exam. The order of administration of the tests was arbitrarily determined to reduce practice or fatigue effects. 23 students took ACCUPLACER first, 23 students took MDTP first, and for 6 the test order is unknown. Within the MDTP Algebra Readiness test, 18 took the CAT first, 17 took PnP first, and 3 are of unknown test order. Within the MDTP Intermediate Algebra test, 5 took the CAT first, 6 took the PnP first and 3 are of unknown test order. Testing order was as balanced as possible within all conditions.

A backup analysis to compensate for the faulty experimental data collection consisted of comparing MDTP Intermediate Algebra scores from Fall 1998 to ACCUPLACER scores from Fall 1999. A subset of 100 scores was to be randomly selected from each semester and sorted by raw score. Scores would have then been matched into pairs with the highest of each being paired together first, then the next highest of each and so on. Demographics of each group would have been compared to check for gross differences in age, gender, ethnicity or other possible factors on influence. Unfortunately, the raw scores for ACCUPLACER for Fall 1999 did not appear to have been entered into the database. While backup data exists, it is not in a readily usable form and has not been converted to usable form by this writing. Due to data this archiving error, the backup analysis could not be performed in a timely fashion.

However, math placement levels were preserved and proportional placements by MDTP were compared to ACCUPLACER. Without raw scores this analysis can only show whether cut scores used for Fall 1999 were consistent with prior placements and cannot suggest what appropriate cut scores for ACCUPLACER should be.

Results

Instructor assessments of student preparedness should show that 75% of students placed by a particular instrument were appropriately prepared to have confidence in the instrument. All math classes except for Math 4, Pre-Calculus, and Math 18, Business Calculus exceeded this level for both ACCUPLACER and MDTP as did the average for all classes combined (Table 1, Table 2, and Table 3). The Math 18 sample only contained 2 students and so has low validity. Of more concern is the Math 4 level of 68% for ACCUPLACER and 67% for MDTP. These groups are not equivalent as the MDTP placements were done exclusively among high school students and ACCUPLACER was used for those entering the Assessment Center at Cabrillo.

CAT and PnP scores for those who took the MDTP Algebra Readiness test were not significantly different between the two testing orders (CAT firsts $z = -0.544$, $p = 0.660$, PnP first $z = -0.314$, $p = 0.757$) nor were they for the MDTP Intermediate Algebra test (CAT first $z = -0.957$, $p = 0.429$, PnP first $z = -0.730$, $p = 0.537$) using the Mann-Whitney U-test.

MDTP Algebra Readiness scores show a significant linear relation with ACCUPLACER scores ($R^2 = 0.291$, $F(1,36) = 14.796$, $p < 0.0005$, Figure 1) as do MDTP Intermediate Algebra scores ($R^2 = 0.419$, $F(1,13) = 9.357$, $p = 0.009$, Figure 2) although graphs of the relations do not appear strictly linear. ACCUPLACER scores showed a significant linear relation to math placement level determined by both MDTP tests ($R^2 = 0.614$, $F(1,50) = 82.093$, $p < 0.005$, Figure 3). These data are suspect as students were allowed to self-select which MDTP they took.

Looking at proportions of placements into the four levels of classes (Arithmetic, Elementary Algebra, Intermediate Algebra, Transfer level) over sequential Fall semesters, we see that Fall 1999, when ACCUPLACER was used, had significantly different proportions from previous semesters (Table 4, Figure 4). While other semesters showed significant deviations as well, the Fall 1999 semester had particularly large adjusted residuals. This suggests that proportions of placements have not remained constant with respect to time. To see if the proportions seen in Fall 1999 were consistent with a pattern of linear increase or decrease, regression models were constructed for each placement level using MDTP data. These models were then used to extrapolate Fall 1999 placements and compared with actual placement proportions (Table 5). While statistical confidence cannot be applied to extrapolations, some of the differences are quite large and do not support the contention that ACCUPLACER placements are consistent with a linear pattern established by 5 years of MDTP placements.

Discussion

While instructor evaluations generally suggest satisfactory placement with ACCUPLACER, the sharp differences seen in placement proportions suggest that cut scores could benefit from adjustment if we desire ACCUPLACER placements to be more consistent with MDTP placements. Based upon differences between observed and expected values, it appears that the current cut scores for ACCUPLACER are too high for the Arithmetic/Elementary Algebra boundary and too low for the Elementary Algebra/Intermediate Algebra boundary and the Intermediate Algebra/Transfer level boundary. Unfortunately, in the absence of ACCUPLACER raw score data, precise recommendations cannot be made.

ACCUPLACER raw scores obtained from the double testing experiment appeared to agree modestly well with MDTP scores. However, enough variability appears in this data to warrant forestalling a firm conclusion without further data collection. Fortunately, data submitted to the State Chancellor's Office have satisfied the validation requirements and we can assume that ACCUPLACER has an acceptable degree of reliability, validity, and lack of bias.

Recommendations

ACCUPLACER is currently approved for use by the Chancellor's Office and therefore is assumed to be effective in placing students. However, there are other factors to consider in deciding upon an assessment tool. How effectively can it be implemented, its data managed, and its operations supported by the vendor? So far, we have experienced serious data management deficiencies with ACCUPLACER from archiving and retrieving scores to incorporating multiple measures. While our current system with MDTP and CAPP to incorporate multiple measures is certainly outdated and could benefit from modernization, it is a system that works. In addition, the College Board has hampered our now moot validation efforts by refusing to release their test item bank forcing us to resort to time consuming methods of content review. Further, their technical support is inadequate being unable to spontaneously answer elementary questions and failing to return calls with correct information. Finally, several requests for copies of their validation studies went unheeded.

In light of these issues, I recommend continuing with MDTP and the CAPP system while continuing to seek an acceptable upgrade of our math assessment process. CAT tests show much promise and should be seriously considered as a replacement to our current paper and pencil tests. An important consideration is the incorporation of multiple measures into a new system. Currently, the CAPP system applies "bonus" points to raw scores based upon examinee responses to background questions. This system was developed and is maintained by an off-campus third party. This has slowed the response time for support and reduces its effectiveness as it is almost solely telephone rather than physical support. A new assessment system should utilize Cabrillo's own Computing Resources personnel to increase the effectiveness and efficiency of implementation and support.

References

California Community Colleges Chancellor's Office. 1998. Matriculation Resource Manual.

Hambleton, R.K. and H. Swaminathan. 1985. Item response theory: Principles and applications. Boston: Kluwer-Nijhoff Publishing.

Lord, F. M. 1980. Applications of item response theory to practical testing problems. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishing.

Wainer, Howard. 1990. Computer adaptive testing: A primer. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishing.

Table 1. Instructor evaluations of student preparedness of those placed with ACCUPLACER in Fall 1999.

Instructor Evaluation	MATH154 El. Alg.		MATH152 Int. Alg.		MATH12 Statistics		MATH13 Finite		MATH18 Bus. Calc		MATH4 Pre-Calc		Total	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%
1 (highest)	45	28	21	21	2	20	1	9	1	50	8	22	78	24
2	49	31	27	27	6	60	2	18		0	12	32	96	30
3	41	26	32	32	1	10	8	73		0	5	14	87	27
4	19	12	15	15	1	10		0	1	50	7	19	43	13
5 (lowest)	6	4	6	6		0		0		0	5	14	17	5
Total	160	100	101	100	10	100	11	100	2	100	37	100	321	100

Table 2. Percent appropriate placement of those placed with ACCUPLACER in Fall 1999.

	% 1,2, or 3	n
MATH154	84	160
MATH152	79	101
MATH12	90	10
MATH13	100	11
MATH18	50	2
MATH4	68	37
Total	81	321

Table 3. Percent appropriate placement of those NOT placed with ACCUPLACER in Fall 1999.

	% 1,2, or 3	n
MATH154	79	66
MATH152	82	28
MATH12	100	2
MATH13	100	8
MATH4	67	6
Total	81	110

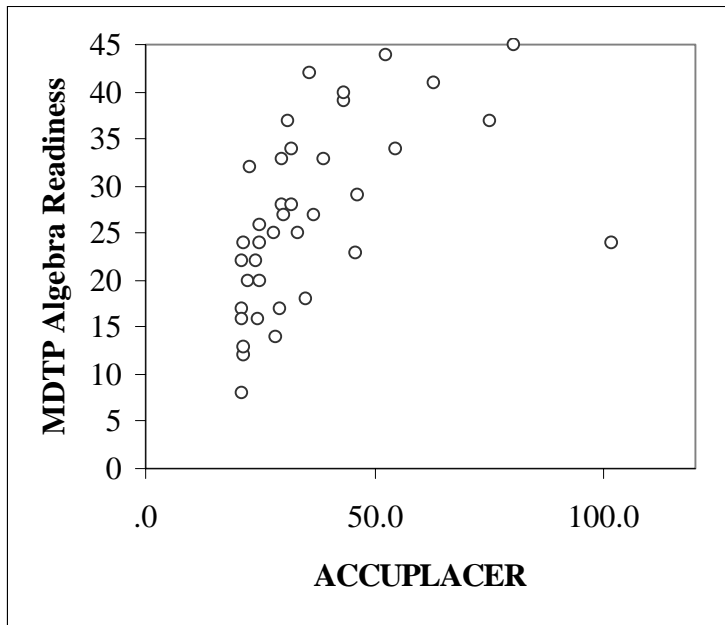


Figure 1. ACCUPLACER algebra scores as predictors of MDTP algebra readiness scores.

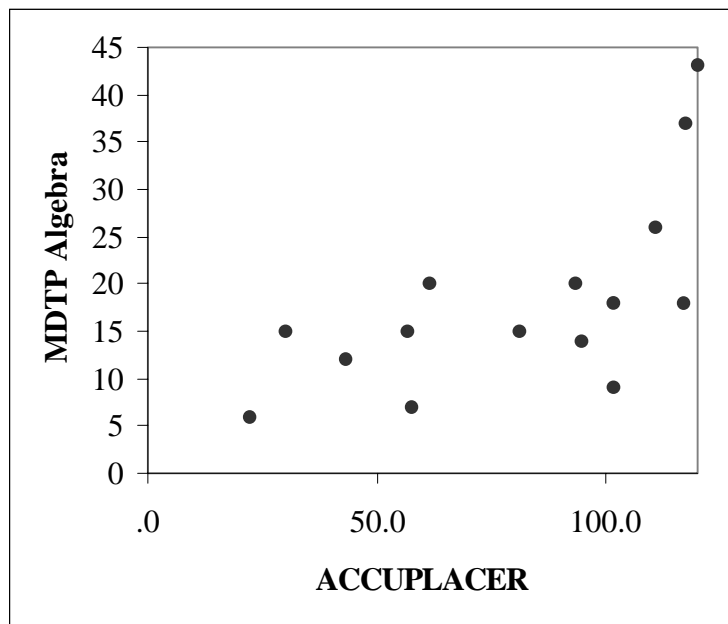


Figure 2 ACCUPLACER algebra scores as predictors of MDTP algebra scores.

Table 4. Number and proportion of math placements using MDTP (1994 – 1998) and ACCUPLACER (1999).

	Frequency	Math 254	Math 154	Math 152	Transfer	Total
Fall 1994	Absolute	751	1046	483	192	2472
	Relative	30.40%	42.30%	19.50%	7.80%	100.00%
	Adj. Res.	-2.3	5	0	-4.4	
Fall 1995	Absolute	693	1051	523	201	2468
	Relative	28.10%	42.60%	21.20%	8.10%	100.00%
	Adj. Res.	-5	5.3	2.3	-3.8	
Fall 1996	Absolute	710	1018	490	225	2443
	Relative	29.10%	41.70%	20.10%	9.20%	100.00%
	Adj. Res.	-3.8	4.2	0.7	-1.8	
Fall 1997	Absolute	675	894	460	237	2266
	Relative	29.80%	39.50%	20.30%	10.50%	100.00%
	Adj. Res.	-2.8	1.7	1	0.4	
Fall 1998	Absolute	832	873	501	289	2495
	Relative	33.30%	35.00%	20.10%	11.60%	100.00%
	Adj. Res.	1.2	-3.3	0.7	2.4	
Fall 1999	Absolute	1043	627	386	344	2400
	Relative	43.50%	26.10%	16.10%	14.30%	100.00%
	Adj. Res.	12.7	-13	-4.7	7.3	
Total	Absolute	4704	5509	2843	1488	14544
	Relative	32.30%	37.90%	19.50%	10.20%	100.00%
Chi-square	344.2295					
df	15					
p	< 0.0005					

Adj. Res. = Adjusted Residual. Similar to a Z score, an adjusted residual less than -2 or greater than +2 is generally considered a large departure from values expected if the proportions of math placements into various classes was equal for all terms.

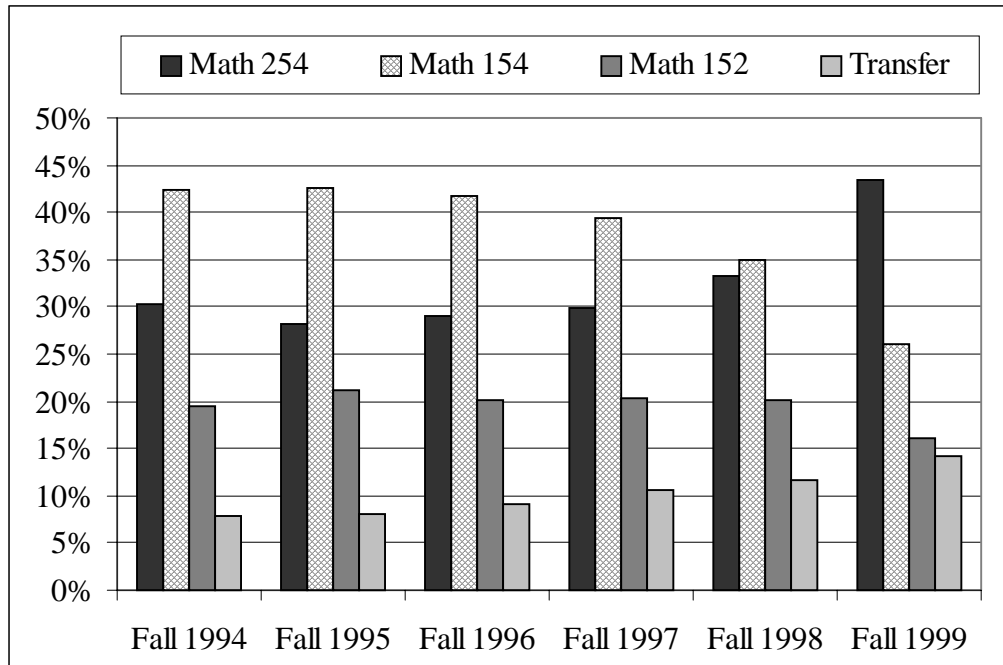


Figure 4. Proportion of math placements with MDTP (1994-1998) and ACCUPLACER (1999).

Table 5. Linearly extrapolated and observed proportions of math placements for Fall 1999.

	linear regression model based on 1994-1998 MDTP	predicted value for Fall 1999 (extrapolated)	observed value for Fall 1999 with ACCUPLACER	difference
Math 254	$Y = 28\% + 1\% * X$	32%	44%	11%
Math 154	$Y = 46\% - 2\% * X$	35%	26%	-9%
Math 152	$Y = 20\% + 0.03\% * X$	20%	16%	-4%
Transfer	$Y = 6\% + 1\% * X$	12%	14%	2%