A Comparison of Conventional and Computer-Based Achievement Testing

Isaac I. Bejar
University of Minnesota

There are at least two types of adaptive achievement testing. In the first type, if there is little variability in achievement at testing time and the purpose of measurement is to classify students into one of two categories (e.g., in mastery testing), then it becomes profitable to adapt the length of the test to the individual. If, on the other hand, there is a fair amount of variability in achievement and the purpose of testing is to assess level of achievement, then in the second type it becomes profitable to adapt both the length *and* difficulty of the test to each individual.

Ferguson (1969) implemented an achievement testing system for the first type which, through the use of decision theory, tested an individual until a decision could be made that the student had reached a prespecified level of achievement. No one has implemented a computer-based achievement testing system for the second type, although there has been an increasing amount of research on computer-based ability measurement.

In implementing an adaptive achievement testing system, serious consideration should be given to the kind of test theory that will serve as the psychometric framework. This is important because it has implications for the creation and calibration of the item pool. The present study is based on item characteristic curve (ICC) theory, which seems to be a more flexible theory compared to theories that could be used in the two types of achievement testing situations described above. Accordingly, this paper is concerned with two questions: (1) Is it possible to construct an item pool for achievement testing based on ICC theory--that is, do ICC models fit observed achievement test data? (2) If the answer is positive, what is the efficiency of an adaptive achievement test compared to a typical paper-and-pencil classroom examination?

The study was done with the cooperation of the Biological Sciences Department at the University of Minnesota. An introductory course was chosen because it had the largest enrollment. The course is offered every quarter, and enrollment ranges from 1,000 to 1,500 per quarter. It is open to all students; both majors and non-majors in the natural sciences enroll. Students are, for the most part, freshmen; but a substantial number of sophomores also enroll. The sexes are about equally represented. According to the course staff, there seem to be no changes in the demographic characteristics of the students from quarter to quarter. Instruction is by means of videotaped lectures shown on closed circuit television and a compulsory laboratory. The course is divided into three units. The first unit covers Chemistry, Energy, and The Cell, in that order; the second unit covers Heredity and Reproduction; and the third unit covers Ecology and Evolution. At the end of each of the first two units a midquarter examination is given. The final examination not only covers these units, but in addition covers the third unit. This paper will be concerned primarily with the first unit.

The Item Pool

Most research to date on adaptive testing has been based on verbal ability (e.g., Lord, 1975; Vale & Weiss, 1975). Typically, verbal items are homogeneous in content; therefore, one of the most important assumptions of latent trait theory, unidimensionality, is justified. By contrast, in achievement testing items are not homogeneous in content. The question arises whether or not a unidimensional model is adequate. There is, of course, no general answer to the question; it must be investigated in every new setting.

Development of the Item Pool

The raw data from which the item pool was formed consisted of answer sheets for the first midquarter exam from a previous academic year. The data matrix for each quarter consisted of between 1,000 and 1,500 respondents and 55 items. Data for each quarter were analyzed separately by Urry's ESTEM program (see Urry, 1976) for estimating item parameters using a minimum chi-square criterion. Table 1 shows the number of items originally available in each of the content areas as well as the number rejected by the program. The program rejects an item if it cannot find a reasonable estimate of one of the three item parameters. This occured with 22% of all items, as seen in Table 1. The remaining 78% of the items presumably fit the three-parameter response model. However, the fact that it was possible to obtain parameters is not in itself evidence of fit. Since the items were from several content areas, it was decided to investigate fit more closely.

Table 1
Summary of Calibration Study

	Chemistry	Cell	Energy	Total
Unique items available[1]	53	60	33	146
Items rejected	16	13	3	32
Percent of items rejected	30	22	9	22
Items calibrated	37	47	30	114

[1] Includes items administered at the final exam.

Dimensionality of the Pool

Correlation of item parameter estimates. It was reasoned that if there were any departures from unidimensionality, they would probably result from content area specific effects. To determine whether or not this was true, for each item a new set of parameter estimates was obtained. These were derived by including only items which belonged to a given content area. The rationale for this method was that if the different content areas measured a single dimension, the item characteristic curve estimated within a content area would be interchangeable with that derived for the entire examination. That is, the regression of content-area-derived parameters on parameters derived from the total examination would have a slope of 1.0 and an intercept of zero.

This was not found to be the case for either the *a* parameter or the *c* parameter. (This may have been attributable to the fact that there were relatively few items in each content area; it is known from simulation studies that a large number of items is required to obtain reliable estimates of the item parameters.) The results for the *b* parameter, on the other hand, showed a great deal of stability.

Table 2 shows the regression statistics for the regression of content-area-based *b* estimates on total test-based estimates for the two first mid-quarter exams. The prediction that the slope would be 1.00 and intercept would be zero was shown to be justified when taking the standard error into account for the first content area in both tests. For the second content area the slope was still 1.00, but the intercept no longer was zero in either midquarter. Finally, for the third content area neither the slope nor the intercept were what they should be. This replicable trend suggests that the metric based on the entire test is not interchangeable with the metric defined within content areas. That is, there is a unique component associated with each content area which was ignored when calibrating all items at once.

Table 2
Regression Statistics for the Regression of the
Content-Area-Based *b* Estimates on the Total Test-Based Estimates

	Slope		Intercept		
	Slope	S.E.	Int.	S.E.	Correlation
Winter					
Chemistry	.94	.03	.00	.03	.99
Cell	1.08	.06	-.41	.09	.98
Energy	.73	.08	.46	.13	.95
Spring					
Chemistry	1.03	.07	.16	.08	.98
Cell	.93	.04	-.31	.06	.99
Energy	.72	.12	.11	.20	.91

Correlation of achievement estimates. The next question was concerned with the extent of the content effect. To answer this question each of the three content areas was scored using the two sets of parameters. It was originally planned to use maximum likelihood scoring but difficulties occurred because there were a large number of cases which did not converge, presumably because some of the content areas had relatively few items. As a result, the data were scored using Owen's (1975) sequential procedure. The inter-content area correlations are seen in Table 3. Both of these matrices could be fitted perfectly by a one-factor model. The maximum likelihood estimates of the loadings on this factor and unique variances are also shown in Table 3.

Table 3
Observed Correlations Among Content Area Scores Using
Content-Area- and Test-Based Item Parameter Estimates

		Content-Area-Based			
Content	Chemistry	Cell	Energy	λ	Uniqueness
Chemistry	.856[1]			.836	.29
Cell	.607	.895		.726	.47
Energy	.511	.444	1.052	.611	.63
		Total Test-Based			
Content	Chemistry	Cell	Energy	λ	Uniqueness
Chemistry	.836[1]			.834	.29
Cell	.623	.937		.747	.44
Energy	.594	.532	.798	.712	.50
Absolute Difference Between Correlations, Loadings, and Uniquenesses					
Chemistry	.020			.002	.00
Cell	.016	.042		.021	.03
Energy	.083	.088	.254	.101	.13

[1] Standard deviations are on the diagonal.

From Table 3 the importance of the content effect can be deduced by computing the difference in unique variances in the two solutions. As seen in Table 3, the estimated unique variances were the same or larger for the content-area-based solution. This is consistent with the earlier hypothesis that there is a unique component associated with performance on each content area beyond that accounted for by general achievement. These differences in unique variance are the proportion of variance attributable to the content area component. In the content areas of Chemistry and Energy this variance was negligible; in the content area of The Cell the variance was not negligible.

Conclusions. These results suggest that in calibrating an achievement test item pool, attention should be given to potential content area influences. It should be pointed out that factor analysis of inter-item correlations is not likely to provide assistance. Such a factor analysis was run and, although a predominant single factor was found, there was no detectable trace of content factors. The regression analysis previously reported appears to be much more powerful. Its usefulness, however, is limited because the analysis can be done only if subsets of items can be identified beforehand as belonging together.

Comparison of the Adaptive and Conventional Tests

Despite the presence of content area effects, each content area could not be calibrated separately to form separate item pools, since there were not enough items available. In effect, the presence of content effects was ignored; and the adaptive testing of achievement proceeded. Although this probably introduced some bias into the results, these scores would not be

Table 4
Values of the *a*, *b*, and *c* Parameters for Items Used in the Conventional Test

Item	*a*	*b*	*c*
3060	.86	-1.31	.29
3067	1.07	-.76	.21
3065	1.17	-1.66	.35
3056	.71	.89	.26
3063	.91	1.51	.35
3073	1.43	-1.57	.31
3058	1.05	-.43	.35
3274	.85	-1.05	.26
3271	.95	1.32	.30
3055	1.71	-.65	.24
3072	1.02	.65	.32
3057	1.20	-1.35	.26
3064	.94	.86	.24
3069	.88	-.01	.35
3054	1.29	-.93	.31
3066	1.05	.53	.31
3268	.97	-.28	.18
3267	1.02	-1.22	.23
3272	1.06	-.81	.35
3070	.95	-1.28	.22
3008	.96	-1.75	.18
3018	1.31	.29	.29
3062	1.47	.43	.30
3061	.95	1.57	.30
3262	.81	.47	.35
3263	.99	2.29	.35
3447	1.18	.93	.32
3443	1.07	-1.64	.35
3438	.70	.21	.27
3448	1.40	.73	.30
3435	.83	-.61	.35
3439	1.36	.64	.32
3436	1.12	1.59	.35
3449	.91	1.26	.14
3440	1.52	2.00	.30
3437	1.95	.66	.28
3427	.92	1.51	.26
3445	1.19	.44	.34
3444	.88	.78	.35

used for grading purposes; therefore,the bias could not affect the individual personally. Comparing modes of administration is often difficult because of the inherent differences of the two testing procedures (cf. Sympson, 1975). Nevertheless, this is a question that must be faced. This study compared the first midquarter Biology examination covering the areas of Chemistry, Energy, and The Cell and a stradaptive test covering the same content areas. The data were collected during fall quarter 1976 and are independent of the data used in the item calibration.

Tests

Classroom test. The classroom exam consisted of the 39 items for which item parameter estimates were available out of the 55 items in the actual test. These 39 items had a mean discrimination of 1.09. The distribution of difficulty was slightly peaked. (It should be pointed out that the criterion used by the Biology staff for assembling the test was a mixture of psychometric, pedagogical, and content considerations.) The item parameters for the 39 items are seen in Table 4.

The stradaptive test. The four major ingredients of the stradaptive strategy are the *item pool, entry point, branching rule,* and *termination criterion* (Weiss, 1973).

The item pool consisted of the 114 items described earlier (see Table 1). The items were assigned to one of nine strata in such a way that there were approximately the same number of items in each stratum. Within stratum the items were placed so that although the content areas were alternated, the most discriminating items were at the top of the stratum. The stradaptive item pool is seen in Table 5.

The entry point for the stradaptive test was determined by the students' reported GPA. For example, if the student reported a GPA of 3.75 or higher, the entry point was at the ninth stratum. At the other extreme, if the student's GPA was less than 2.00, the entry point was the first stratum (i.e., the easiest stratum).

The branching rule used in the present study was to present an item from the next more difficult stratum following a correct answer and an item from the next less difficult stratum following an incorrect answer. After responding to the first item in the entry stratum, the student was given the first item from the next lower stratum if the answer was incorrect or the first item from the next higher stratum if the answer was correct. Thereafter, the student was branched to the next unadministered item in the next higher or lower stratum,depending on whether the answer was correct or incorrect. The exception to this rule occurred if the testee was at the most difficult stratum. In that case,after a correct answer the next item in that same stratum was given. Similarly, fôr the student in the least difficult stratum, an incorrect response led to the next item in that stratum.

Table 5
Values of the *a*, *b*, and *c* Parameters for Items in the Stradaptive Test by Stratum

Item	*a*	*b*	*c*	Item	*a*	*b*	*c*	Item	*a*	*b*	*c*
Stratum 9:(most difficult):				Stratum 6:				Stratum 3:			
3209	2.77	2.29	.29	3047	1.66	.44	.29	3021	1.96	-.49	.21
3417	2.67	3.02	.35	3213	.93	.52	.35	3217	1.06	-.48	.14
3033	1.54	2.44	.35	3041	1.51	.23	.35	3038	1.71	-.93	.21
3251	2.60	2.39	.35	3405	1.40	.55	.32	3215	1.59	-.82	.23
3406	1.31	2.48	.35	3218	.82	.58	.12	3011	1.32	-.86	.20
3045	1.02	2.48	.29	3019	1.31	.29	.29	3216	1.27	-.62	.18
3242	.94	2.40	.35	3207	.70	.46	.28	3221	1.25	-.52	.17
3407	1.02	2.41	.29	3431	.70	.28	.20	3049	1.15	-.71	.18
3241	.91	2.09	.13	3000	1.24	.52	.35	3255	1.14	-.72	.26
3414	.88	2.29	.32	3046	1.18	.24	.22	3246	1.10	-.72	.28
3402	.83	2.44	.35	3042	1.15	.37	.27	3022	1.01	-.48	.30
3247	.82	2.42	.35	3050	1.13	.35	.18	3017	.99	-.58	.16
3228	.67	2.49	.31	3034	1.01	.37	.28	3224	.80	-.50	.27
Stratum 8:				Stratum 5:				Stratum 2:			
3409	4.68	1.28	.00	3220	1.79	-.03	.26	3023	2.40	-1.15	.35
3234	3.54	1.73	.00	3005	1.43	.11	.35	3202	1.81	-.99	.21
3018	.89	1.25	.35	3425	1.36	.17	.23	3415	.85	-.96	.35
3204	1.14	1.66	.35	3039	1.12	.12	.34	3245	1.34	-.96	.21
3422	1.47	1.50	.35	3214	1.12	.03	.23	3236	1.26	-1.20	.33
3411	1.36	1.23	.35	3412	1.12	.19	.35	3020	1.23	-1.28	.17
3250	.91	1.94	.29	3051	1.29	.21	.28	3028	1.12	-1.26	.35
3206	.74	1.51	.21	3403	.99	.18	.19	3226	1.09	-.98	.20
3410	1.30	1.34	.31	3211	.88	.01	.13	3210	1.04	-1.22	.35
3429	1.25	1.24	.28	3002	.82	.13	.14	3239	1.04	-1.13	.21
3419	1.23	1.48	.25	3426	.68	.07	.22	3013	1.00	-.97	.39
3421	1.17	1.15	.35	3423	.66	.16	.27	3257	.98	-1.02	.25
3427	.92	1.51	.26					3036	.92	-1.18	.16
3420	.68	1.62	.35	Stratum 4:				3014	.86	-1.24	.14
								3238	.82	-1.06	.21
Stratum 7:				3256	2.31	-.33	.26	3032	.77	-1.06	.27
3408	2.51	1.05	.31	3430	1.15	-.30	.29				
3258	1.24	.81	.35	3031	1.47	-.33	.35	Stratum 1:			
3432	1.72	.67	.35	3254	2.28	-.17	.27	3027	1.67	-1.38	.35
3048	1.35	.66	.33	3237	1.54	-.37	.18	3249	.91	-1.69	.17
3413	1.40	.76	.35	3404	.65	-.29	.35	3428	.90	-1.56	.35
3219	1.23	.62	.21	3244	1.35	-.44	.23	3205	1.25	-1.53	.19
3035	.90	.68	.28	3240	.98	-.28	.15	3235	1.15	-1.40	.28
3433	1.35	.86	.30	3208	.76	-.16	.12	3029	1.13	-1.50	.28
3230	.90	.87	.35	3006	.77	-.37	.33	3201	1.07	-1.34	.23
3012	.75	.80	.38	3259	.69	-.41	.20	3008	.96	-1.75	.18
3260	.71	.84	.28					3252	.79	-1.77	.35
								3003	.96	-1.76	.34
								3044	.87	-1.42	.15

The termination rule used in this study was that testing was stopped if in any stratum a student had answered 5 items and 20% or more of them had been incorrect answers, or if 50 items had been administered, whichever occurred first.

Scoring. Both the adaptive and classroom data were scored by the method of maximum likelihood, using the Newton-Raphson numerical procedure with a set of locally written programs. For scoring purposes the item parameter estimates were edited so that the maximum discrimination was 2.50, the maximum absolute value of the difficulty parameter was set to 3.00, and the maximum "guessing" parameter was set to .35. Less than 1% of the ability estimates failed to converge during scoring.

Criteria for Comparison

One of the most important contributions of latent trait theory to psychometrics has been the concept of information. Unlike reliability and related concepts, information is a local measure of the accuracy of estimation of a testee's achievement levels.

Samejima (1969) defines the test information function, in general, as

$$I(\theta) = E\,\frac{\partial^2 L_v(\theta)}{\partial\theta^2}\quad, \qquad [1]$$

where $L_v(\theta)$ is the log-likelihood function of the response vector v. Thus, test information is the expected value of the second derivative of the log-likelihood function. This apparently arbitrary quantity is useful because its reciprocal, $1/I(\theta)$, is the minimum sampling variance of an estimator. As such, it is a measure of the best that could be accomplished in estimating with a given response model if an appropriate scoring procedure is used. Maximum likelihood estimation provides, asymptotically, estimators with that property.

Since it is an expected value, information is, in a sense, a prediction of the model; and in fact, it does not depend on a response vector. This is a useful property when making theoretical comparisons. For empirical comparisons, however, it seems more appropriate to base the result on a statistic closer to the data. That statistic may be called the *observed* information (cf. Edwards, 1972). Samejima (1973) has referred to observed information as the *response vector information function.* Equations 2 and 3 give, respectively, the expressions for the test information function and the response vector information function.

$$I(\theta) = \sum_g D^2 a_g^2\ \Psi[DL_g(\theta)] - P_g(\theta) D^2 a_g^2\ \Psi[DL_g(\theta) - log\ c_g] \qquad [2]$$

$$\hat{I}(\theta) = \sum_g D^2 a_g^2\ \Psi[DL_g(\theta)] - u_g(\theta) D^2 a_g^2\ \Psi[DL_g(\theta) - log\ c_g] \qquad [3]$$

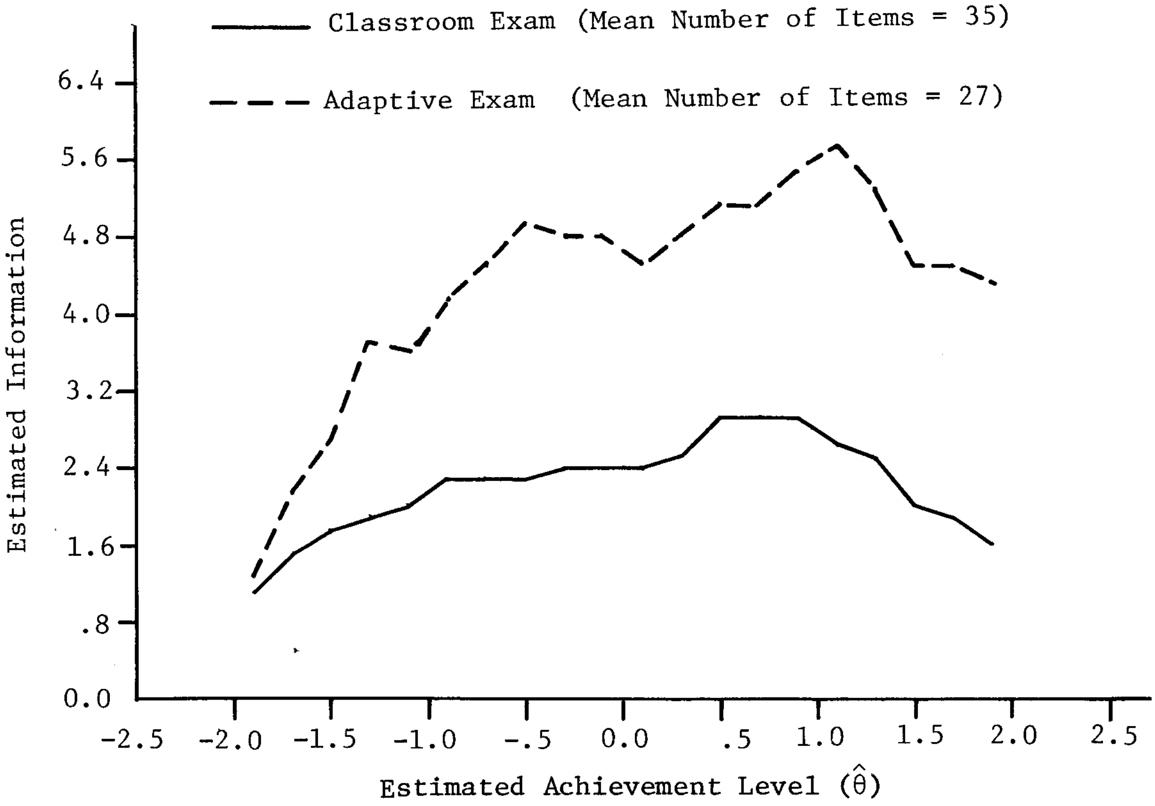
Equations 2 and 3 are identical with the exception that the second term on the right is weighted by $P(\theta)$ (i.e., the probability of passing the item) in one case and by u_g(=1 if the answer is right, 0 otherwise) in the other. To the extent that $P(\theta)$ is an accurate estimate of the probability that u_g=1, these two kinds of information functions will differ very little.

Results

Approximately 350 students came to the laboratory for testing between one day and three weeks after the classroom exam. For their participation, they received points toward their final course grade and a computer-printed report of the questions they had taken as part of the adaptive test.

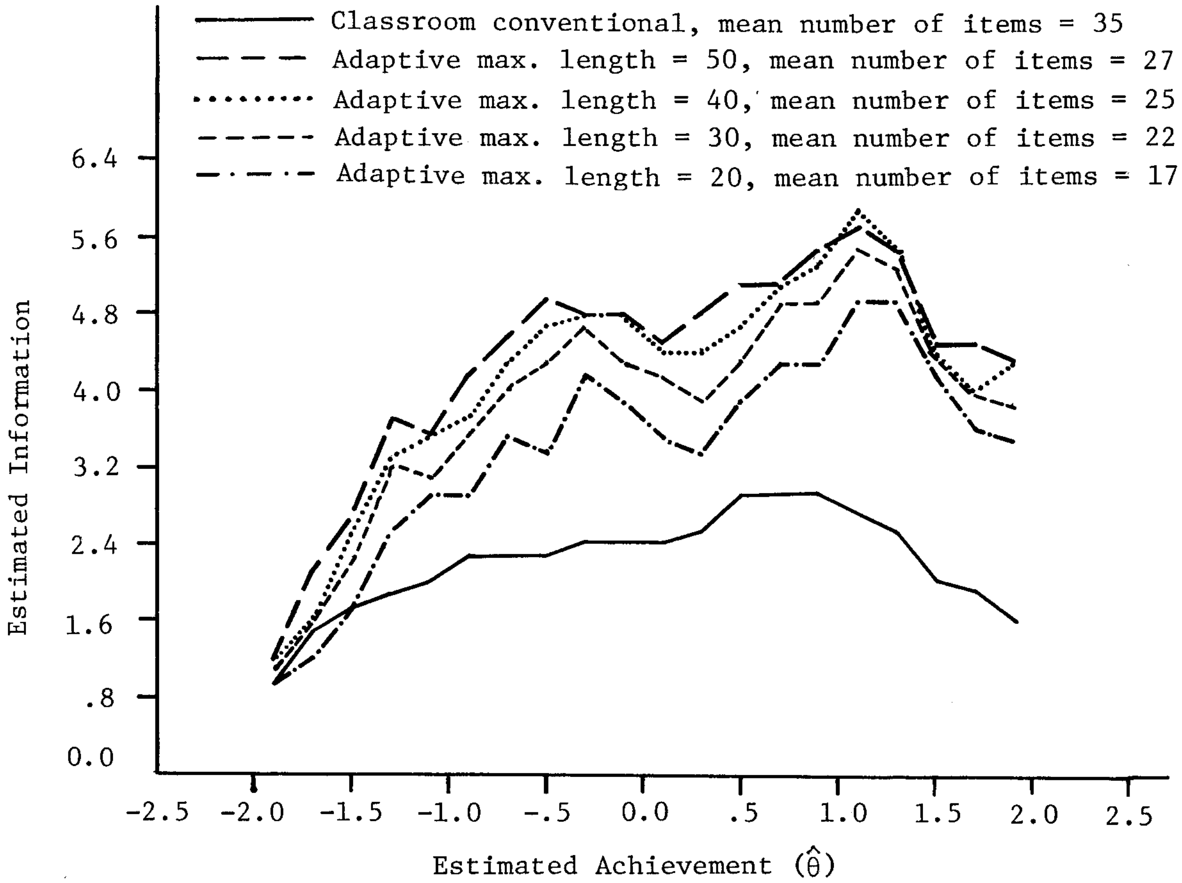
Adaptive vs. classroom test. Figure 1 shows the response vector information functions for the adaptive and classroom exams. To obtain the curves the maximum likelihood estimates of θ between -2.00 and +2.00 were divided into intervals of .20. The mean observed information for testees in a given interval was assigned to the midpoint of that interval. These values are plotted in Figure 1. The adaptive test is far superior; however, several factors must be considered to put this in perspective.

Figure 1
Response Vector Information Curves for
Classroom Conventional Test and Adaptive Test



As indicated earlier, the stradaptive test had variable test length. The mean test length in the 20 intervals of θ was fairly constant. The mean test length across all students was 27 items; for the conventional test the mean was 35. Despite the fact that the stradaptive test was shorter, on the average, it yielded more information. To see in retrospect what the results would have been with a shorter adaptive test, the adaptive data were rescored reducing the maximum test length from 50 to 40, 30 and 20 items. The mean number of items for the maximum lengths of 40, 30, and 20 were respectively 25, 22, and 17 items. The results are seen in Figure 2. Note that the adaptive test with a maximum test length of 20 still yielded a substantially higher amount of information compared to the classroom exam, while shortening the test considerably.

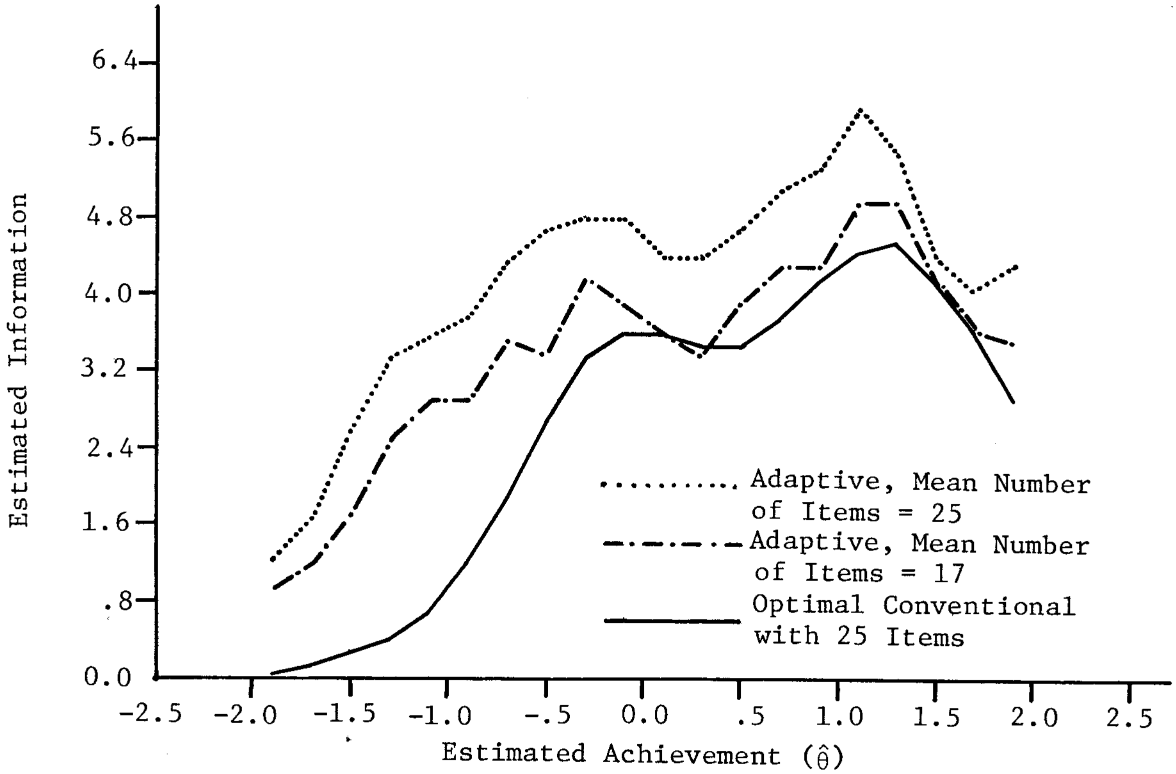
Figure 2
Response Vector Information Curves for
Classroom Conventional Test and Adaptive Test at Four Test Lengths



Adaptive vs. an ideal conventional test. These results are not surprising because the conventional test used for comparison was not designed to be optimal. A fairer comparison might have been to contrast the adaptive test against an ideal conventional test. For the ideal conventional test the top items were chosen from each of the seven most difficult strata until a maximum of 25 items was reached. This resulted in a conventional test with mean a=1.70 and approximately rectangularly distributed b's in the interval -1.00 to 3.00. The test information function of the conventional test and the response vector information function for the adaptive test with a maximum length of 20 and 40 items are seen in Figure 3.

The information for the optimal conventional test was very low for θ's below -1.00. This was a function of the distribution of item difficulties chosen for the test. It was reasoned that an optimal achievement test need not have very high information in the lower end of θ. In addition, by concentrating the difficulty of the items in a restricted range of θ, the conventional method was given a better chance against the adaptive test. The maximum information of this optimal conventional test was 4.59 at θ=1.10. The adaptive information functions also peaked at θ=1.10;and the shortest adaptive test yielded 10% more information with 30% fewer items on the average. Compared to the information for the optimal conventional test, the information for the adaptive test with a maximum length of 40 items (which resulted in a mean of 25 items per student, i.e., the same number of items as in the optimal conventional test) resulted in an even larger increase in information.

Figure 3
Information Curves for Optimal Conventional Test and Adaptive Test at Two Test Lengths



Summary and Conclusions

Two questions have been addressed in this paper. The first question was whether it is possible to construct an item pool based on the ICC model which could be used in adaptive achievement testing. In general, the answer was found to be positive. The spread of difficulties and discrimination in the item pool were such that adaptive testing would be effective. However, unique response components associated with the different content areas were also identified. The magnitude of these components was not large,but must have introduced certain biases into the comparison of the two testing procedures. As a result of ignoring the influence of content-specific factors on test performance, the parameters of the estimated item characteristic curves derived from the entire test may have included a mis-specification bias. Since the adaptive testing procedure relied on these possibly misestimated item parameters for the sequential selection of items, it is likely that the advantage of adaptive testing over conventional achievement testing was underestimated.

The second question addressed was how effective adaptive achievement testing is compared to conventional testing. The answer was that adaptive testing can drastically reduce testing time while yielding more precise scores than an actual conventional or an ideal conventional test. Although the answer is gratifying, it is one which should be expected from theoretical studies. In fact, the stage is rapidly being approached in which the increased efficiency of adaptive testing is no longer an issue. Future research in adaptive achievement testing should concern itself with the truly unique needs of achievement testing. Two such needs are the ability to perform multi-content branching and the need to assess growth. Work on multi-content branching is already under way (Urry, 1977; Weiss & Brown, 1977). Little, however, seems to have been done in the areas of the assessment of growth by means of computerized testing. Hopefully, that gap will be filled in the near future.

References

Brown, J., & Weiss, D. J. An adaptive testing strategy for achievement test batteries (Research Report 77-6). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1977.

Edwards, A. W. F. Likelihood: An account of the statistical concept of likelihood and its application to scientific inference. London: Cambridge University Press, 1972.

Ferguson, N. L. The development, implementation, and evaluation of a computer-assisted branched test for a program of individually prescribed instruction. Unpublished doctoral dissertation, University of Pittsburgh, 1969.

Lord, F. M. A broad-range tailored test of verbal ability. In C. L. Clark (Ed.), Proceedings of the First Conference on Computerized Adaptive Testing. Washington, DC: U.S. Civil Service Commission, Bureau of Policies and Standards, 1975.

Owen, R. J. A Bayesian sequential procedure for quantal response in the context of adaptive mental testing. Journal of the American Statistical Association, 1975, 70, 351-356.

Samejima, F. Estimation of latent ability using a response pattern of graded scores. Psychometrika, 1969, Monograph Supplement No. 17.

Samejima, F. A comment on Birnbaum's three-parameter logistic model in the latent trait theory. Psychometrika, 1973, 38, 221-234.

Sympson, J. B. Evaluating the results of computerized adaptive testing. In D. J. Weiss (Ed.), Computerized adaptive trait measurement: Problems and prospects (Research Report 75-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1975. (NTIS No. AD A018675)

Urry, V. W. A five-year quest: Is computerized adaptive testing feasible? In C. L. Clark (Ed.), Proceedings of the first conference on computerized adaptive testing. Washington, DC: U.S. Civil Service Commission, Bureau of Policy and Standards, 1976.

Urry, V. W. A multivariate model-sampling procedure and a method of multidimensional tailored testing. Paper presented at Computerized Adaptive Testing '77 Conference, Minneapolis, July, 1977.

Vale, C. D., & Weiss, D. J. A study of computer-administered stradaptive ability testing (Research Report 75-4). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1975. (NTIS No. AD A013185)

Weiss, D. J. The stratified adaptive computerized ability test (Research Report 73-3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1973. (NTIS No. AD 768376)

Acknowledgments

This research was supported by funds from the Navy Personnel Research and Development Center, Army Research Institute, Air Force Human Resources Laboratory, Defense Advanced Research Projects Agency, and the Office of Naval Research, monitored under contract No. N00014-76-C-0627 NR150-389 with the Office of Naval Research. The assistance of Kathleen A. Gialluca and G. Gage Kingsbury in data analysis is deeply appreciated, as is the cooperation of Kathy Swart and Norman Kerr of the General Biology staff in the College of Biological Sciences at the University of Minnesota,in providing access to the students and data.